G-quadruplexes
G-quadruplexes

Lead Guest Editor: Shozeb Haider
Guest Editors: Gary Parkinson and Tom Marsh
Editorial Board

Ashis Basu, USA
Ben Berkhout, Netherlands
Emery H. Bresnick, USA
Jyoti Chattopadhyaya, Sweden
Grigory Dianov, UK

Joachim Engels, Germany
Ramon Eritja, Spain
Tapas Hazra, USA
Hyouta Himeno, Japan
Shigenori Iwai, Japan

Xingguo Liang, China
Luis A. Marky, USA
Satoshi Obika, Japan
Tatiana S. Oretskaya, Russia
Hiroshi Sugiyama, Japan
Contents

G-Quadruplexes (GQU)
Shozeb Haider , Gary N. Parkinson, and Thomas C. Marsh

Electrochemical and AFM Characterization of G-Quadruplex Electrochemical Biosensors and Applications
Ana-Maria Chiorcea-Paquim , Ramon Eritja , and Ana Maria Oliveira-Brett
Review Article (20 pages), Article ID 5307106, Volume 2018 (2018)

Putative HIV and SIV G-Quadruplex Sequences in Coding and Noncoding Regions Can Form G-Quadruplexes
Petra Krafčíková, Erika Demkovičová, Andrea Halaganová, and Viktor Víglaský

Telomeric G-Quadruplexes: From Human to Tetrahymena Repeats
Erika Demkovičová, Ľuboš Bauer, Petra Krafčíková, Katarína Tlučková, Petra Tóthova, Andrea Halaganová, Eva Valušová, and Viktor Víglaský
Research Article (14 pages), Article ID 9170371, Volume 2017 (2018)

On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids
Ewan K. S. McRae, Evan P. Booy, Gay Pauline Padilla-Meier, and Sean A. McKenna
Review Article (11 pages), Article ID 9675348, Volume 2017 (2018)

On the Helical Structure of Guanosine 5’-Monophosphate Formed at pH 5: Is It Left- or Right-Handed?
Gang Wu, Irene C. M. Kwan, Zhimin Yan, Yining Huang, and Eric Ye
Research Article (7 pages), Article ID 6798759, Volume 2017 (2018)

In What Ways Do Synthetic Nucleotides and Natural Base Lesions Alter the Structural Stability of G-Quadruplex Nucleic Acids?
Janos Sagi
Review Article (45 pages), Article ID 1641845, Volume 2017 (2018)
G-Quadruplexes (GQU)

Shozeb Haider,1 Gary N. Parkinson,1 and Thomas C. Marsh2

1UCL School of Pharmacy, University College London, London, UK
2Department of Chemistry, University of St. Thomas, St. Paul, MN, USA

Correspondence should be addressed to Shozeb Haider; shozeb.haider@ucl.ac.uk

Received 4 March 2018; Accepted 7 March 2018; Published 30 April 2018

The articles in this special edition provide a view of the complexity that guanine bases and their modifications present in both lower and complex organisms and eukaryotic organisms. Guanine nucleotides and guanine-rich nucleic acids have a well-known propensity to self-associate into highly stable structures with a common structural motif of four Hoogsteen H-bonded, coplanar guanines referred to as a G-quartet (aka G-tetrad). A relatively distinctive feature of stacked G-quartets compared to Watson-Crick base paired structures is the presence of a coordination pocket that is often occupied by a variety of physiological cations such as $K^+$, $\text{NH}_4^+$, $\text{Ca}^{2+}$, and $\text{Na}^+$ as well as others such as $\text{Sr}^{2+}$ and $\text{Pb}^{2+}$. Overall, the stability and morphology of G-quartet structures are influenced by a combination of intrinsic and external factors (e.g., sequence and coordinating cation) with a growing degree of predictability. Evidence of the broader role G-quadruplexes play in information metabolism has also grown tremendously in the past decades giving us potential therapeutic targets for cancer and similar diseases. Furthermore, the key interactions that guanine makes have been used for their application as electrochemical biosensors and development of nanoscale devices.

Two articles are presented by V. Viglasky et al. In one article, they scan the human and simian immunodeficiency provirus genome for putative G-quadruplex forming sequences. They rationalize that targeting the G-quadruplexes in HIV offers an attractive therapeutic target, which would be of particular use in the development of novel antiviral therapies. The analysis of G-rich regions can provide researchers with a path to find specific targets that could be of interest for specific types of virus. In the other article, they trace the effect of modifications in telomeric sequences in Tetrahymena and human repeats. The human telomeric and protozoan telomeric sequences differ only in one purine base (TTAGGG to TTGGGG). They go on to demonstrate that while the substitution does not affect the formation of G-quadruplexes, it does result in an alteration of topology. The results also show that the stability of the substituted derivatives increased in sequences with greater number of substitutions. This observation is somewhat analogous to the phenomenon observed in human telomeric sequences, where the same TTAGGG sequence can adopt six different types of topologies depending upon the chemical environment.

The importance of G-quadruplex function can be emphasized just based on the number of proteins that are being reported that have an association with them. While G-quadruplexes are being touted as potential drug targets and small molecule compounds are being developed, still very little is known about how these multistranded nucleic acids structures interact with proteins. The review by S. A. McKenna et al. focuses on the recognition and comparison of G-quadruplexes by proteins and small peptides, mainly taking into account the X-ray crystallographic and NMR structures, as well as biochemical investigations of binding specificity. These structural features can be used to study and rationally design molecules that target protein-G-quadruplex interactions.

In another article, J. Sagi reviews the structural stability of natural base lesion and synthetic nucleotides. The comprehensive review on the thermodynamic stability of the modified G-quadruplex folds and the stereochemical preferences of more than 70 synthetic and natural derivatives of nucleotides substituting for the natural ones determine the stability and their conformation. The stability of the nucleotide analogs depends on the glycosidic bond conformation, their position of occurrence, and the quadruplex...
fold. Base modifications hold extreme importance in epigenetic studies. An insight into the thermodynamics of G-quadruplex structural stability can be useful in engineering a stable G-quadruplex topology and in exploring the action of base modifications on G-quadruplex architectures both in vitro and in vivo.

In another article, G. Wu et al. explore a five-decade old question, what is the handedness of 5'-GMP helical structure? They report, using NMR and IR spectroscopy, the structural details of the helix, which contains 15 nucleotides per 4 turns and only C3'-endo sugar puckering. The switch in pH from 8 to 5 results in the helix being devoid of Na⁺ ions, which is in sharp contrast to the 5'-GMP helix formed at pH 8 where the central channel is filled with Na⁺ ions.

In another article, A. M. Oliveira-Brett et al. review the recent advances in the applications of G-quadruplexes as biosensors. G-quadruplex electrochemical biosensors have received particular attention, since the electrochemical response is particularly sensitive to the DNA structural changes from a single-stranded, double-stranded, or hairpin into a G-quadruplex configuration. The development of an increased number of G-quadruplex aptamers, which combine the G-quadruplex stiffness and self-assembling versatility with the aptamer high specificity of binding to a variety of molecular targets, allowed the construction of biosensors with increased selectivity and sensitivity. The electrochemical characterization, design, and applications of G-quadruplex electrochemical biosensors in the evaluation of metal ions, G-quadruplex ligands, and other small organic molecules, proteins, and cells are reviewed. The electrochemical and atomic force microscopy characterization of G-quadruplexes is presented. Different G-quadruplex electrochemical biosensors design strategies, based on the DNA folding into a G-quadruplex, the use of G-quadruplex aptamers, or the use of G-quadruplex DNAzymes, are discussed.

Acknowledgments

Finally, the editors of this issue would like to thank all the authors and referees involved in the peer-reviewing process.

Shozeb Haider
Gary N. Parkinson
Thomas C. Marsh
Guanine-rich DNA sequences are able to form G-quadruplexes, being involved in important biological processes and representing smart self-assembling nanomaterials that are increasingly used in DNA nanotechnology and biosensor technology. G-quadruplex electrochemical biosensors have received particular attention, since the electrochemical response is particularly sensitive to the DNA structural changes from single-stranded, double-stranded, or hairpin into a G-quadruplex configuration. Furthermore, the development of an increased number of G-quadruplex aptamers that combine the G-quadruplex stiffness and self-assembling versatility with the aptamer high specificity of binding to a variety of molecular targets allowed the construction of biosensors with increased selectivity and sensitivity. This review discusses the recent advances on the electrochemical characterization, design, and applications of G-quadruplex electrochemical biosensors in the evaluation of metal ions, G-quadruplex ligands, and other small organic molecules, proteins, and cells. The electrochemical and atomic force microscopy characterization of G-quadruplexes is presented. The incubation time and cations concentration dependence in controlling the G-quadruplex folding, stability, and nanostructures formation at carbon electrodes are discussed. Different G-quadruplex electrochemical biosensors design strategies, based on the DNA folding into a G-quadruplex, the use of G-quadruplex aptamers, or the use of hemin/G-quadruplex DNAzymes, are revisited.

1. Introduction

In addition to its genetic role, DNA represents one of the most important and smart self-assembling nanomaterials, being largely used in DNA nanotechnology and biosensor technology [1]. A DNA-electrochemical biosensor is a sensing device composed of a DNA layer (the biological recognition element) immobilized on the electrode surface (the electrochemical transducer), to detect target analytes that interact with DNA at nanoscale. The analytes will induce morphological, structural, and electrochemical changes in the DNA layer, which are further translated into an electrochemical signal, Scheme 1 [2–9]. The DNA-electrochemical biosensors are very robust, easy to miniaturise, present excellent detection limits, use small analyte volumes, and have the ability to be used in turbid biofluids, which make them exceptional tools for rapid and simple on-field detection. They also represent good models for simulating nucleic acid interactions with cell membranes, specific DNA sequences, proteins, pharmaceutical drugs, and hazard compounds [2–11].

The DNA is composed of nucleotides, each containing a phosphate group, a sugar group, a nitrogen base, the purines adenine (A) and guanine (G), and the pyrimidines thymine (T) and cytosine (C), Scheme 2(a). The main structural conformation for natural DNA is the double-stranded DNA in Watson-Crick base pairs, Scheme 2(b), the cellular DNA being almost exclusively in this form [12]. However, DNA can be found in a variety of other conformations, such as double-helices with different types of loops (bulge, internal, hairpin, junction, knotted loops, etc.), single-strands, triple-helices, or four-stranded secondary structures (e.g., i-motifs and G-quadruplexes (GQs)) [11–13].
The GQs are four-stranded secondary structures, Scheme 2(c), formed by planar associations of four G bases, named G-quartets, held together by eight Hoogsteen hydrogen bonds, Scheme 2(b). The G-quartets are stacked on top of each other and stabilized by π–π hydrophobic interactions. Monovalent cations, such as K⁺ and Na⁺, are coordinated to the lone pairs of electrons of Oπ in each G.

The GQ structures are polymorphic, and a variety of topologies have been observed by nuclear magnetic resonance (NMR) or crystallography, either as native structures or complexed with small molecules [14–17].

According to the number of strands, GQs can be classified as monomers (unimolecular, intramolecular, e.g., the human telomeric DNA d[AG₃₂(T₃AG)₂] in the presence of K⁺ ions, Protein Data Bank (PDB) entry 1KFI [18]), dimers (bimolecular, intermolecular, e.g., the Oxytricha nova telomeric sequence d(G₅T₄G₅) in the presence of K⁺ ions, PDB entry 1JPQ [19]), or tetramers (tetramolecular, intermolecular, e.g., the Tetrahymena telomeric sequence d(TG₄T) in the presence of Na⁺ and Ca²⁺ ions, PDB entry 2GW0 [20]), Scheme 2(c).

According to the strand polarity (i.e., the relative arrangement of adjacent strands), the GQs present parallel or antiparallel orientations, according to the glycosidic torsion angle, they present anti or syn orientation, and according to the orientation of the connecting loops, they can be lateral, diagonal, or both [21–24].

The GQ sequences are found in chromosomess’ telomeric regions, oncogene promoter sequences, RNA 5′-untranslated regions (5′-UTR), and other relevant genome regions, where they may influence the gene metabolism process and also participate in DNA replication, transcriptional regulation, and genome stability [14, 21–32]. The GQ formation has been associated with a number of diseases, such as cancer, HIV, diabetes, and aging [14, 23]. They are also considered important cancer-specific molecular targets for anticancer drugs, since the GQ stabilization by small organic molecules can lead to telomerase inhibition and telomere dysfunction in cancer cells [22, 33, 34].

Due to GQs biological role, extraordinary stiffness, and the ability to self-organize in more complex two-dimensional networks and long nanowires, they have become relevant in structural biology, medicinal chemistry, supramolecular chemistry, nanotechnology, and biosensor technology [14, 22, 23, 25, 35–37].

Short chain G-rich DNA sequences that form GQ structures are now used as recognition elements in GQ electrochemical biosensor devices, since the electrochemical response is particularly sensitive to the DNA sequence structural variations from a single-stranded, double-stranded, or hairpin configuration into a GQ configuration. In addition, short aptamers able to form GQs received a great deal of attention, since they are highly specific in binding to small molecules, proteins, nucleic acids, and even cells and tissues. These GQ aptamers combine the G-quadruplex stiffness and self-assembling versatility with the aptamer high specificity of binding, which allowed the construction of GQ electrochemical biosensors with increased selectivity and sensitivity.

The recent advances on the characterization of the G-rich DNA sequences, at the surface of electrochemical transducers, and the design and applications of GQ electrochemical biosensors for the detection of metal ions, GQ ligands, and other small organic molecules, proteins, and cells will be presented.

2. Electrochemical and AFM Characterization of G-Quadruplexes

Understanding the redox behaviour and adsorption process of the DNA probe at electrochemical transducers is critical for the design and successful application of DNA-electrochemical biosensors [2–7, 9–11, 38–46]. At carbon electrodes, the voltammetric studies showed that DNA bases, Figure 1(a), nucleosides, and nucleotides are all electroactive [47–49]. The double- and single-stranded DNA oxidation in solution shows two anodic peaks, corresponding to the G residues oxidation (G₃) and A residues oxidation (A₃), Figure 1(b) [2–7, 9–11, 38–46]. The T and C residues oxidation is more difficult to detect, since it occurs with a very low current, at very high positive potentials, near the potential of oxygen evolution.

The electrochemical behaviour of G-rich DNA sequences, able to self-assemble into GQ configurations, Scheme 2(c), was recently studied [54–59].

2.1. Short Chain G-Rich Oligonucleotides. The first report on the electrochemical oxidation of GQs concerned the investigation of two different length thrombin-binding aptamer (TBA) sequences, d(G₃T₅G₃TGTG₃T₃G₅) and d(G₃T₅G₃-TGTG₃T₃G₅) [54]. Both TBA sequences form chain-like unimolecular GQs in K⁺ ion containing solutions, consisting of two G-quartets connected by two TT loops and a single TGT loop. The different adsorption patterns and degree of surface coverage observed by atomic force microscopy (AFM) were correlated with the sequence base composition, presence/absence of K⁺ ions, and voltammetric behaviour observed by differential pulse (DP) voltammetry.

In the absence of K⁺, in Na⁺ containing solutions, the formation of GQs was very slow. The oxidation of both TBA sequences showed only one anodic peak, corresponding to the G residues oxidation in the TBA single-strands [54]. In the presence of K⁺, both TBA sequences folded into GQs. Since no adsorption of the stable and rigid GQs occurred,
Scheme 2: (a) Chemical structure of DNA nucleotides, nucleosides, and bases, (b) Watson-Crick base-pairing and G-quartet, and (c) G-quadruplex configurations. [Adapted from [11, 13] with permission.]
only a few single-stranded sequences were observed on the highly oriented pyrolytic graphite (HOPG) surface. DP voltammetry showed the decrease of the G residues oxidation peak and the occurrence of a new GQ peak at a higher potential, corresponding to the G residues oxidation in the GQs. The GQ higher oxidation potential was due to the greater difficulty of electron transfer from the inside of the GQ structure to the electrode surface compared to the electron transfer from the G residues in the more flexible single-strands.

The redox behaviour and adsorption of the d(G)_{10}, d(TG)_{9}, and d(TG_T) oligodeoxynucleotide (ODN) sequences were studied by AFM and DP voltammetry at carbon electrodes [50, 56, 57]. All sequences fold into parallel tetramolecular GQ structures, Scheme 2(c)-right. The results demonstrated that GQ formation was directly influenced by the ODN sequence and concentration, pH, and presence of monovalent cations, Na⁺ versus K⁺, Figures 2(c) and 2(d). DP voltammetry allowed the detection of the ODN single-strands folding into GQs and G-based nanostructures, Figure 2(a), in freshly prepared solutions, at concentrations 10 times lower than usually detected using other techniques generally employed to study the GQ formation.

Single-stranded d(G)_{10}, d(TG)_{9}, and d(TG_T) ODNs, in Na⁺ containing solutions and for short incubation times, were detected using AFM as network films and polymeric structures, Figure 2(b), and using DP voltammetry by the occurrence of only the G residues oxidation peak (G r), Figure 2(c) [13, 50]. GQ structures, in Na⁺ containing solutions and for long incubation times, or in K⁺ containing solutions, were detected using AFM as spherical aggregates, Figure 2(b) (white arrows). DP voltammetry showed the decrease of the G residues oxidation peak and the GQ oxidation peak occurrence, increase, and shift to more positive potentials, in a time-dependent manner, Figure 2(c) [13, 50]. Concerning the self-assembling into higher-order nanostructures, the homo-ODN sequence d(G)_{10} was the only sequence forming G-nanowires observed using AFM, Figure 2(b) (black arrows), d(TG)_{9} formed short G-based superstructures that
Figure 2: (a) Representation of d(G)_{10}-GQ and GQ-based nanowire; (b) AFM image of 0.3 μM d(G)_{10} in pH 7.0, 100 mM K⁺ ions, at 24 h incubation; (c, d) DP voltammograms baseline corrected in 3.0 μM d(G)_{10} in pH 7.0: (c) incubation time dependence and (d) K⁺ ions concentration dependence at 0 h incubation. [Adapted from [50] with permission.]
were adsorbed as rod-like shape aggregates, and d(TG₄T) formed no nanostructures, due to the presence of T residues at both 5' and 3' ends [50].

The d(TG₄T) telomeric repeat sequence of the free-living ciliate protozoa Tetrahymena forms tetramolecular GQs that are considered simpler models of biologically relevant human quadruplexes, being used to obtain high resolution data on pharmaceutical drug-DNA interactions [51]. The well-known conformation of the d(TG₄T) GQ and its extraordinary stiffness enabled the d(TG₄T) to be considered a good building block candidate for the development of novel devices, with medical and nanotechnology applications. The d(TG₄T) self-assembling from single-strand into GQ, influenced by the Na⁺ versus K⁺ ions concentration, was successfully detected using AFM on HOPG, Figure 3(a), and DP voltammetry at glassy carbon (GC) electrode, Figures 3(b) and 3(c) [51]. The d(TG₄T) GQs self-assembled very fast in K⁺ and slowly in Na⁺ containing solutions, revealing a time and a K⁺ ions concentration dependent adsorption process and redox behaviour, Figure 3(c).

AFM images of d(TG₄T) spontaneously adsorbed from freshly prepared solutions (0h incubation), in the presence of Na⁺ ions, showed only randomly oriented polymeric structures of 0.89 ± 0.1 nm height, due to the adsorption of single-stranded d(TG₄T) molecules, as shown in the high amplification image from Figure 3(a)-up-left.

AFM images after 48 h incubation showed three adsorption morphologies: randomly oriented polymeric structures and network films of 0.81 ± 0.1 nm height, due to the adsorption of d(TG₄T) single-strands, spherical aggregates of 2.15 ± 0.6 nm height, due to the adsorption of short tetramolecular d(TG₄T) GQs, and sporadically short nanowires of 0.80 ± 0.1 nm height and length up to 100 nm. In order to be able to distinguish the presence of the three morphologies, a larger amplification image, Figure 3(a)-up-middle, has been chosen.

AFM images after 7 days' incubation also showed three adsorption morphologies: very rarely, randomly oriented polymeric structures and network films of 0.81 ± 0.1 nm height, due to the adsorption of d(TG₄T) single-strands, spherical aggregates of 2.05 ± 0.5 nm height, due to the adsorption of d(TG₄T) GQs, and oriented polymeric domains of 0.81 ± 0.1 nm height, adsorbed along one of the three axes of symmetry of the HOPG basal planes. Again, in order to distinguish the three adsorption morphologies, a larger amplification image, Figure 3(a)-up-right, has been presented.

AFM images of d(TG₄T) immediately after the addition of K⁺ ions (0 h incubation) showed two adsorption morphologies: randomly oriented polymeric structures of 0.71 ± 0.2 nm height, due to the adsorption of d(TG₄T) single-strands, and spherical aggregates of 1.87 ± 0.4 nm height, due to the adsorption of d(TG₄T) GQs, Figure 3(a)-down-left. Increasing the incubation time to 48 h (larger magnification image from Figure 3(a)-down-middle) and 7 days' incubation (Figure 3(a)-down-right), the number of 1.85 ± 0.5 nm height aggregates increased, while the 0.80 ± 0.1 nm height polymeric domains decreased.

The optimum K⁺ ions concentration for the formation of d(TG₄T)-GQs was similar to the healthy cells intracellular K⁺ ions concentration. The d(TG₄T) higher-order nanostructures self-assembled slowly in Na⁺ ion solutions and were detected by AFM as short nanowires and nanostructured films, Figure 3(a)-up. The absence of higher-order nanostructures in K⁺ ion solutions, Figure 3(a)-down, showed that the rapid formation of stable GQs induced by the K⁺ ions is relevant for the good function of cells.

In another report, the discrimination between double-stranded and GQ DNA was achieved at the gold electrode surface [60], based on the selective interaction between a [Ru(NH₃)₆]³⁺ redox label and DNA sequences able to form GQs with different folding type and numbers of G-quartets, d(G₂T₄G₂TGTG₂T₂G₂), d(T₄G₂T₂G₄TGTG₂T₂G₂), d(AG₂T₄- -AG₂T₄AG₂T₄AG₂T₄), d(GTAG₂T₄G₂TGTG₂T₂G₂), and d(T₂CGTCGT₂T₂G₂CAG₂T₂G₂TGACT). A characteristic voltammetric peak was observed, due to a strong association between the [Ru(NH₃)₆]³⁺ redox label and the GQs, which was not detected for double-stranded DNA sequences.

2.2. Long Chain G-Rich Polynucleotides. Long chain polynucleotides poly(dG) and poly(G) are widely prevalent in the human and other genomes at both DNA and RNA levels and are used in DNA-electrochemical biosensors, as models to determine the preferential interaction of drugs with G-rich segments of DNA.

AFM at HOPG, Figures 4(a)–4(c), and DP voltammetric studies at GC electrode, Figure 4(d), showed that, in the presence of monovalent Na⁺ or K⁺ ions, the poly(G) single-strands self-assembled into short GQ regions for short incubation times. Large poly(G) GQ aggregates with low adsorption were formed after long incubation times, Figures 4(e)–4(g) [52]. DP voltamograms in freshly prepared poly(G) solutions showed only the G residues oxidation peak (G₂), Figure 4(d), due to the G residues oxidation in the poly(G) single-strands.

Increasing the incubation time, the G residues oxidation peak decreased and disappeared, and a GQ oxidation peak in the poly(G) GQ morphology appeared, at a higher oxidation potential, dependent on the incubation time. The GQ oxidation peak current presented a maximum after 10 days' incubation and reached a steady value after ~17 days' incubation, Figure 4(d).

3. G-Quadruplex Electrochemical Biosensor Applications

Several electrochemical strategies are used in DNA-electrochemical biosensors applications:

(i) The direct label-free detection of DNA bases electrochemical current, monitoring the modifications of the G and A₇ oxidation peaks

(ii) The detection of redox reactions of reporter label molecules

(iii) The detection of charge transport reactions mediated by the π-π interaction between DNA stacked bases.
Figure 3: (a) AFM image of 0.3 µM d(TG₄T) in pH 7.0, in the presence of Na⁺ and K⁺ ions, at different incubation times; (b) representation of the d(TG₄T) single-strand and GQ electrochemical detection; (c) DP voltammograms baseline corrected in 3.0 µM d(TG₄T) in pH 7.0: incubation time and K⁺ ions concentration dependence. [Adapted from [51] with permission.]
Figure 4: (a–c) AFM images of 5 μg mL⁻¹ poly(G) in pH 7.0, in the presence of K⁺ ions, at (a) 0 h, (b) 24 h, and (c) 21 days' incubation; (d) DP voltammograms baseline corrected in 100 μg mL⁻¹ poly(G) in pH 7.0, in the presence of K⁺ ions, at (red dashed line) 0 h, (black solid line) 24 h, (black dashed line) 10 days, and (red solid line) 21 days' incubation and control (black dashed-dotted line) 3 μM d(G)₁₀ at 24 h incubation; (e–g) representation of poly(G) adsorption process: (e) poly(G) single-strand, (f) poly(G) single-strand with short GQ regions, and (g) poly(G) single-strand with larger GQ regions. [Reproduced from [52] with permission.]
However, for the design of GQ electrochemical biosensors, the use of redox molecular labels for electrochemical current amplification has been preferred. The GQ electrochemical biosensor applications, for the detection of metal ions, GQ ligands and other small organic molecules, proteins, and cells, will be discussed.

3.1. Metal Ions

3.1.1. Folding-Based G-Quadruplex Electrochemical Biosensor. Besides the G-rich DNA sequence requirements, the coordination of monovalent cations is essential for the GQ formation and stability [61]. The physiologically relevant cations for GQ formation are the K\(^{+}\) and Na\(^{+}\), but, under specific conditions, cations such as Rb\(^{+}\), Cs\(^{+}\), NH\(_4\)^{+}, Tl\(^{+}\), Sr\(^{2+}\), and Ba\(^{2+}\), and Pb\(^{2+}\) also influence the GQ formation. Based on the G-rich DNA conformational change from a single-strand to a GQ in the presence of metal ions, Scheme 3, different folding-based GQ electrochemical biosensors have been reported [62, 63].

A GQ electrochemical biosensor for K\(^+\) ions, based on the d(G\(_3\)T\(_3\)G\(_3\)TGTG\(_2\)T\(_3\)G\(_2\)) sequence switching from single-strand into a GQ in the presence of K\(^+\), was developed [64]. The ODN structural modifications were detected via changes on the electron transfer between a ferrocene (Fc) redox label and the gold electrode surface. Similar approaches were used for d(T\(_4\)G\(_3\)T\(_2\)AG\(_3\)T\(_2\)AG\(_3\)T\(_2\)AG\(_3\)) [65], d(T\(_2\)CACATACTGACTCAG\(_3\)) [66], and d(T\(_3\)G\(_2\)T\(_2\)-G\(_3\)TGTG\(_2\)T\(_2\)G\(_2\)) [67], self-assembled layers in the presence of Fc, either by voltammetry or electrochemical impedance spectroscopy (EIS). Folding-based GQ electrochemical biosensors for monitoring the Tb\(^{3+}\) ions at d(T\(_{20}\)G\(_3\)(T\(_2\)AG\(_3\))\(_3\)) [68] and Pb\(^{2+}\) at d(T\(_4\)C\(_2\)A\(_2\)CG\(_3\)T\(_2\)-G\(_2\)TGTG\(_2\)T\(_2\)G\(_2\)) [69] and modified gold electrodes, in the presence of Fc redox labels, were also described.

Small electroactive GQ ligands, able to intercalate into the GQ structure, such as ethyl green [62] and crystal violet [63], were also successfully employed as redox labels. Ethyl green GQ intercalator was applied for the development of a GQ electrochemical biosensor for the determination of Pb\(^{2+}\) ions [62], at the surfaces of carbon paste and multiwalled carbon nanotube paste electrodes. The electrochemical determination of Pb\(^{2+}\) was achieved by following the structural modification of the d(G\(_3\)T\(_4\)) sequence from a single-strand into a GQ in the presence of Pb\(^{2+}\), followed by the ethyl green intercalation into the GQ, which caused changes in the ethyl green reduction peak current. In another example, a GQ electrochemical biosensor for Pb\(^{2+}\) detection was based on the electrochemical current of the crystal violet GQ ligand, with the d(G\(_3\)T\(_4\)) sequence immobilized at a gold electrode surface [63].

3.1.2. Hemin/GQ DNAzyme Electrochemical Biosensor. Hemin/GQ DNAzyme electrochemical biosensors represent one of the most popular building assays of GQ electrochemical biosensors [70]. In peroxidase hemin/GQ DNAzyme, the complex formed by hemin, an iron-containing porphyrin, with GQ DNA sequences, leads to an improved peroxidase activity of hemin and facilitates a redox reaction between a target molecule (the substrate, e.g., 3,3',5,5'-tetramethyldibenzidine, hydroquinone, or ferrocene methyl alcohol) and H\(_2\)O\(_2\). The target molecule oxidation product is electrochemically detected, Scheme 4.

A hemin/GQ DNAzyme electrochemical biosensor for the detection of Hg\(^{2+}\) was developed [71], based on a bifunctional ODN sequence that contained a Hg-specific domain and a GQ domain, immobilized on gold electrode surface. The interaction between the GQ domain and hemin generated a hemin/GQ complex, which catalyzed the electrochemical reduction of H\(_2\)O\(_2\), producing amplified readout currents for Hg\(^{2+}\) interaction events.

In a more complex design, the hemin/GQ DNAzyme was used to develop a surface plasmon resonance and electrochemical biosensor for Pb\(^{2+}\) ions [72]. A complex consisting of the Pb\(^{2+}\)-dependent DNAzyme sequence and a ribonuclease-containing nucleic acid sequence (corresponding to the substrate of the DNAzyme) linked to a G-rich ODN sequence was assembled on gold electrode surfaces. In the presence of Pb\(^{2+}\) ions, the Pb\(^{2+}\)-dependent DNAzyme cleaved the substrate, leading to the separation of the complex and
to the self-assembly of the hemin/GQ complex. The electrochemical detection of Pb\(^{2+}\) showed a detection limit of 1 pM and a good selectivity.

In a different approach, taking advantage of the hemin/GQ ability to act both as a NADH oxidase, assisting the oxidation of NADH to NAD\(^+\) together with the generation of H\(_2\)O\(_2\) in the presence of dissolved O\(_2\), and a peroxidase DNAzyme to biosensitize the reduction of the produced H\(_2\)O\(_2\), a hemin/GQ DNAzyme electrochemical biosensor for Hg\(^{2+}\) detection was developed [73]. The sensor showed improved detection limit and excellent selectivity against other interfering metal ions.

3.2. G-Quadruplex Ligands and Other Small Organic Compounds

3.2.1. Folding-Based G-Quadruplex Electrochemical Biosensor.
The GQ ligands are small molecules that bind to G-rich DNA sequences, which are considered novel therapeutic targets for anticancer drug development. They have the ability to induce and stabilize the DNA folding into GQ configurations at the level of telomeres, preventing the telomeric DNA from unwinding and opening to telomerase and thus indirectly targeting the telomerase and inhibiting its catalytic activity, which further leads to the senescence and apoptosis of tumour cells.

Recently, remarkable progress has been made in the development of selective GQ ligands that entered in clinical trials for cancer therapy, presenting significant telomerase inhibition or suppression of the transcription activity of oncogenes. Examples may include the trisubstituted acridine compound BRACO-19 and, more recently, a series of new triazole-linked acridine ligands, for example, GL15 and GL7, with enhanced selectivity for human telomeric GQ binding versus duplex DNA binding.

BRACO-19, GL15, and GL7 present complex, pH-dependent, and adsorption-controlled irreversible oxidation mechanisms, at GC electrode [74, 75]. The interaction between DNA and the acridine ligands GL15 and GL7 was investigated in incubated solutions and using DNA-, poly(G)-, and poly(A)-electrochemical biosensors [75]. Both GL15 and GL7 interacted with DNA in a time-dependent manner, with preferential affinity for the G-rich segments, but did not cause DNA oxidative damage.

The interactions of the GQ-targeting triazole-linked acridine ligand GL15 with the short chain length Tetrahymena telomeric DNA repeat sequence d(TG\(_4\)T) and with the long poly(G) sequence have been studied [53]. The results showed that GL15 interacts with both sequences in a time-dependent manner. In the presence of GL15, GQ formation was detected by AFM via the adsorption of GL15-d(TG\(_4\)T) GQ and GL15-poly(G) GQ small spherical aggregates and large GL15-poly(G) GQ assemblies and by DP voltammetry via GL15 and G\(_4\) oxidation peak current decrease and disappearance and the occurrence of a GQ oxidation peak, Figure 5.

The GL15 interaction with d(TG\(_4\)T) and poly(G) was directly influenced by the presence in solution of monovalent Na\(^+\) ions, Figure 5(a), or K\(^+\) ions, Figure 5(b). These results were consistent with the interaction of triazole-linked acridine derivatives with terminal G-quartets in individual GQs, Figure 5-right. The binding, in Na\(^+\) or K\(^+\) ions solutions, of GL15 to d(TG\(_4\)T) and poly(G) strongly stabilized the GQs and accelerated GQ formation, although only the K\(^+\) ions containing solution promoted the formation of perfectly aligned tetramolecular GQs [53].

An electrochemical biosensor for the investigation of GQ ligands telomerase inhibitors prepared by modification of a GC electrode with gold nanoparticles and GQ d((G\(_3\)T\(_2\)A\(_3\)G\(_3\)) and i-motif d((C\(_3\)T\(_2\))\(_2\)C\(_3\)) DNA sequences was developed [76]. EIS results showed the increase of the charge transfer resistance with increasing the GQ ligand concentration, due to the GQ ligand interaction. 1,4-Dihydropyridine derivatives showed good GQ affinity in the concentration range from 5.0 to 700 μM, and the ligands selectivity was studied using different control double-stranded DNA sequences.

A GQ electrochemical biosensor for the detection of GQ ligands ethidium bromide and polyamines spermine or spermidine was developed [77], which was based on the immobilization of the 30-mer d(G\(_3\)A\(_3\)G\(_3\)A\(_5\)G\(_3\))ATGGC sequence on a pretreated multilayered carbon nanotubes modified GC electrode. The characteristics of the modified electrode and the GQ interaction with ethidium bromide and polyamines spermine or spermidine were investigated, via the [Ru(NH\(_3\))\(_3\)\(^{3+}\) peak current decrease with increasing ligands concentration.

Apart for GQ ligands, GQ electrochemical biosensors for the detection of other small organic molecules were developed, using aptamers with specific recognition for the target analyte. Several GQ aptamers for small organic molecules have been described in the literature, for example, hematoporphyrin IX, hemin, ochratoxin, and ATP [78], and some were already used for the development of GQ electrochemical biosensors.

Ochratoxin A (OTA) is a secondary metabolite of fungi strains like Aspergillus ochraceus, Aspergillus carbonarius, and Penicillium verrucosum that can contaminate a large number of food supplies. An impedimetric GQ electrochemical biosensor for the detection of OTA was developed, based on the OTA specific aptamer sequence d(GATCG\(_3\)GTG\(_3\)TA\(_3\)G\(_3\)AGCATCG\(_2\)ACA) [79]. The aptamer was covalently immobilized onto a mixed Langmuir–Blodgett monolayer composed of polyaniline-stearic acid and deposited on indium tin oxide coated glass plates. The sensor showed a 0.24 nM detection limit for OTA [79]. This methodology was improved, a detection limit of 0.12 nM for OTA was achieved, and the biosensor has been also successfully applied for OTA determination in food samples [80].

Another design for the OTA detection proposed the use of a long polyethylene glycol spacer chain, which led to the formation of long tunnels at the surface of screen printed carbon electrodes, with the OTA aptamers acting as gates. The OTA specific binding to the aptamer led to changes in the aptamer configuration and, consequently, to a peak current decrease [81].

In a different approach, OTA was detected at a GQ electrochemical biosensor that used a hairpin OTA aptamer and
site-specific DNA cleavage of the TaqAl restriction endonuclease and a streptavidin-horseradish peroxidase label [82].

3.2.2. Hemin/GQ DNAzyme Electrochemical Biosensor. A hemin/GQ DNAzyme electrochemical biosensor for the detection of the GQ ligands 5,10,15,20-tetra-(N-methyl-4-pyridyl)porphyrin (TMPyP4) and N,N-bis[2-(1-piperidino)-ethyl]-3,4,9,10-perylenetetracarboxylic diimide (PIPER) [83] was described. The biosensor was prepared using the human telomeric DNA sequence d(AG₃(T₂AG₃)₃) immobilized at the pyrolytic graphite electrode surface. Both the hemin and the GQ ligand bound simultaneously to the GQ structure, and neither PIPER nor TMPyP4 destroyed the hemin/GQ complex. Voltammetric and spectrometric methods were simultaneously employed to verify the interactions and binding stoichiometry between the GQ ligands and the hemin/GQ complex. The binding stoichiometry was determined to be 2:1 for TMPyP4-hemin/GQ and 4:1 for PIPER-hemin/GQ.

A common strategy for hemin/GQ DNAzyme electrochemical biosensors for the detection of small organic compounds that do not directly bind to the GQ consisted in the modification of the electrode surface by an ODN sequence that contains two domains. One domain is capable of forming a GQ structure, which binds the hemin and is used as amplification strategy. The other is an aptamer domain able to specifically bind the analyte, which may form or not a GQ structure. In the presence of the analyte and hemin, the hemin/GQ structures were formed on the electrode surface, while the analyte protein was bound to the aptamer part. This strategy was successfully used to design a hemin/GQ DNAzyme electrochemical biosensor for the detection of adenosine monophosphate (AMP) [84]. Upon interaction of a hairpin ODN sequence with AMP, the AMP-aptamer complex was formed, leading to the hairpin opening and the
formation of the hemin/GQ complex. On the other hand, the adenosine deaminase was able to convert the AMP substrate into inosine monophosphate that lacked the affinity for the aptamer sequence. The sensor achieved a the 1 μM detection limit for AMP [84].

Adenosine triphosphate (ATP) was detected at a hemin/GQ DNAzyme electrochemical biosensor that used two aptamers for ATP and for hemin recognition [85]. Two ODN sequences were designed, the first one immobilized on a gold electrode surface and containing the ATP aptamer and a part of the hemin aptamer and the second one containing the complementary strand of ATP aptamer and the rest of hemin aptamer. In the presence of ATP, the duplex between the two ODN sequences opened, the second ODN sequence diffused into the solution, and the hemin/GQ DNAzyme electrochemical current disappeared.

A dual-functional electrochemical biosensor for ATP and H2O2 from cancer cells was developed based on a hemin/G-quadruplex DNAzyme [86]. The double-stranded conformation of the ATP aptamer, immobilized on gold electrodes, changed upon ATP binding, forming a stable GQ, and a hemin/G-quadruplex DNAzyme, after addition of hemin, was formed. The electrochemical current of the Fc redox label increased in the presence of both ATP and H2O2.

A hemin/G4 DNAzyme based impedimetric biosensor was used to detect the environmental metabolite 2-hydroxyfluorene (2-HOFlu) [87]. Using the hemin/G4 peroxidase activity to catalyze the oxidation of 2-HOFlu by H2O2, the sensor achieved a 2-HOFlu detection limit of 1.2 nM in water and 3.6 nM in spiked lake water samples. The assay was also selective over other fluorene derivatives.

Hemin/GQ DNAzyme electrochemical biosensors for the detection of other organic compounds, such as the toxin microcystin-LR [88] and the pollutant agent naphthol [89], were also described.

3.3. Proteins. Due to aptamers’ high selectivity, sensitivity, and reliability, the electrochemical biosensors are also very attractive tools for protein detection. TBA was the first aptamer observed in NMR studies to fold into a GQ structure [90], and it was shown that GQ formation is critical for thrombin specific recognition. TBA remains nowadays the most widely used aptamer in GQ electrochemical biosensors research. However, GQ aptamers that bind specifically to other protein targets have been selected, for example, nucleolin, signal transducer and activator of transcription STAT3, human RNase H1, protein tyrosine phosphatase Shp2, VEGF, HIV-1 integrase, HIV-1 reverse transcriptase, HIV-1 reverse transcriptase, HIV-1 nucleocapsid protein, M. tuberculosis polyphosphate kinase 2, sclerostin, and insulin [78].

3.3.1. Folding-Based G-Quadruplex Electrochemical Biosensor. Many GQ electrochemical biosensors for the detection of proteins are based on the aptamer structural modifications in the presence of the analyte, from a single-, double-strand, or hairpin configuration, into a GQ configuration. Generally, they consist in an aptamer modified with a redox label immobilized on the electrode surface, while the analyte is present in solution [91, 92], Scheme 5.

The first folding-based GQ electrochemical biosensor for thrombin was prepared by covalently attaching methylene blue (MB) labelled TBA sequences d(TA2G6-CATCT-C4G2T2G2TGTG2T2G2T) [93] and d(T2G2C2A2CG2T2-G2TGTG2T2G2) [94] to gold electrode surfaces. In the absence of thrombin, the immobilized TBA sequences remained unfolded, allowing the MB label to be in close proximity to the electrode surface and electron transfer occurred. Upon thrombin binding, the electron transfer between the MB redox label and the gold surface was stopped, due to the GQ aptamer formation. Similar GQ electrochemical biosensors for thrombin were developed, using the TBA sequence d(G5T5G2TGTG2T5G2) labelled with Fc [95–97], and FeO2-nanoparticles [98].

In another approach, the folding-based GQ electrochemical biosensor for thrombin was prepared by covalently attaching to a gold electrode an ODN sequence that contained a 15-base TBA sequence at its 3'-end and formed a double-helix with a MB-tagged partially complementary ODN sequence [99]. In the presence of thrombin, the 15-base TBA sequence self-assembled into a GQ, releasing the 5’ end of the MB-tagged ODN sequence as a flexible, single-stranded element and thus producing a detectable current. This strategy achieved an increase in current of ~300% with a saturated thrombin target, but the GQ electrochemical biosensor was not reusable. In a similar procedure, a TBA sequence d(C5G2A2CG2T2G2TGTG2G2) labelled with Fc immobilized on gold electrode was also employed [100].

A folding-based GQ electrochemical biosensor for thrombin was developed based on thrombin-induced split aptamer fragments conjunction [101]. The 15-base TBA sequence was split into two fragments, the d(A5G2T2G2TG) sequence that was attached to a gold electrode and the d(TG2T2G2T6) sequence modified with a Fc redox label. The thrombin-induced association of the two fragments increased the concentration of Fc at the gold surface, which was monitored by voltammetry.

A label-free impedimetric folding-based GQ electrochemical biosensor for thrombin was developed based on an electropolymerized poly(pyrrrole-nitrioltriacetic acid) film onto the surface of a platinum electrode, followed by complexation of Cu2+ ions and immobilization of histidine-TBA.
sequences \( d(G_2T_2G_2TGTG_2T_2G_2) \) \[102\]. The biosensor presented high sensitivity for the detection and quantification of thrombin via EIS detection, without a labelling step.

Different electrode surface modification strategies have been used to improve the sensitivity of the GQ electrochemical biosensors, and examples may include the use of gold disk microelectrode arrays \[103\], gold electrode surface modified by polyamidoamine (PAMAM) dendrimer \[104\], GC electrode surface modified by gold nanoparticles \[105\] and multiwalled carbon nanotubes \[106\], magnetic nanobeads \[107\] and quantum dots-coated silica nanospheres \[108\], and gold nanoparticles \[109\] at graphite electrodes.

3.3.2. Sandwich-Type GQ Electrochemical Biosensor. Many aptamers recognize specifically different positions on the analyte, as in the case of TBA that recognizes both the fibrinogen and heparin binding sites of thrombin, a property that was used for the development of sandwich-type GQ electrochemical biosensors.

The first sandwich-type GQ electrochemical biosensor reported in the literature for thrombin detection presented an aptamer-analyte-aptamer format, Scheme 6(a), and led to the selective detection of 1 \( \mu \)M of thrombin \[110\]. The sensor was built up by two aptamer layers, TBA 1 immobilized onto the gold electrode surface and used for capturing the thrombin analyte, and TBA 2, labelled with glucose dehydrogenase and used for the electrochemical detection, Scheme 6(a). The TBA sequences 15-mer \( d(G_2T_2G_2TGTG_2T_2G_2) \) and 29-mer \( d(AGTC_2GTG_2TAG_3CAG_2T_2G_2TGACT) \) were used as either TBA 1 or TBA 2, and similar results were obtained. In a very similar methodology \[111\], the 29-mer TBA was labelled with pyrroloquinoline quinone (PQQ) redox cofactor glucose dehydrogenase (GDH), which led to a current increase due to the electroactive product generated by the enzyme reaction and allowed the selective detection of more than 10 nM of thrombin.

The limit of detection of sandwich-type GQ electrochemical biosensor with aptamer-analyte-aptamer format was further lowered by employing different types of redox labels on the TBA 2, such as peroxidase \[112\], platinum nanoparticles \[113\], gold nanoparticles \[114\], and cadmium sulphide quantum dots \[115, 116\]. Using a more complex design, based on conductive graphene-3,4,9,10-perylenetetracarboxylic dianhydride nanocomposites as sensor platform, and PtCo nanochains-thionine-Pt-horseradish peroxidase labelled secondary TBA sequences \( d(G_2T_2G_2TGTG_2T_2G_2) \) for current amplification, a \( 6.5 \times 10^{-16} \) M detection limit for thrombin was achieved \[117\].

A sandwich-type GQ electrochemical biosensor for thrombin detection with an antibody-analyte-antibody format, Scheme 6(b), was also described \[118\]. The thrombin analyte was immobilized directly onto a nanogold-chitosan composite modified GC electrode surface, via a polyclone antibody. A 58-mer TBA sequence \( d(GACAGACGTGT-TGCAGACTACTG_2T_2G_2T_2G_2G_2TAGTCAGACACACTGCTGTG) \) labelled with MB was used for detection, and a 0.5 nM detection limit for thrombin was obtained \[118\].

In another approach, a sandwich-type GQ electrochemical biosensor for thrombin detection with an aptamer-analyte-antibody format, Scheme 6(c), was applied \[119\]. The sensor was built up by immobilizing the thrombin analyte to gold nanoparticles-doped conducting polymer nanorods via TBA \( d(G_2T_2G_2TGTG_2T_2G_2) \) sequences, the detection being performed with the Fc redox labelled antibody. The
electrocatalytic oxidation of ascorbic acid by the Fc redox label allowed a low detection limit of 0.14 pM for thrombin. The biosensor was successfully tested in a real human serum sample for the detection of spiked concentrations of thrombin.

3.3.3. GQ DNAzyme Electrochemical Biosensor

(a) Hemin/GQ DNAzyme. Based on the peroxidase hemin/GQs DNAzyme characteristics, Scheme 4, different hemin/GQ DNAzyme electrochemical biosensors were described. A hemin/GQ DNAzyme electrochemical biosensor, based on a TBA d(G2T2G2TGTG2T2G2) sequence, was applied for the thrombin detection [120]. In the presence of thrombin, the hemin/GQ DNAzyme activity increased, providing the amplified electrochemical readout currents, the sensor exhibiting high sensitivity and selectivity.

In a more complicated strategy, the thrombin detection was achieved using a dual current amplification scheme [121]. Gold nanoparticles were first electrodeposited onto single wall nanotube graphene modified electrode surface, for the immobilization of electrochemical probe of nickel hexacyanoferrates nanoparticles. Subsequently, another gold nanoparticles layer was electrodeposited for further immobilization of TBA sequences, which later formed the hemin/GQ DNAzyme. On the basis of the dual amplifying action, a detection limit of 2 pM for thrombin was obtained.

A hemin/GQ DNAzyme electrochemical biosensor for thrombin detection, based on background noise reduction by exonuclease I (Exo I), was also described [122]. The TBA sequences were self-assembled onto gold nanoparticles modified screen printed carbon electrode surfaces. In the absence of the target thrombin, the TBA sequences were digested by Exo I, which impeded the hemin association, significantly reducing the background current noise. The thrombin binding stabilized the TBA GQ and prevented it from being degraded by Exo I, the hemin/GQ complex formation generating the amplified electrochemical current. The introduction of Exo I significantly enhanced the current to noise ratio of the electrochemical sensor response.

An electrochemical aptasensor for thrombin that used the cocrystallization of hemin/GQ DNAzyme and octahedral Cu3O-Au nanocomposites for signal amplification was designed [123]. Gold nanoparticles were grown directly on the surface of the octahedral Cu3O nanocrystals. The Cu3O-Au nanocomposites obtained were simultaneously used for signal amplifying molecules and as nanocarriers. The hemin/GQ DNAzyme was formed by intercalating hemin into the d(G2T2G2TGTG2T2G2) TBA sequences and the electrocatalytic toluidine blue and was immobilized onto the Cu3O-Au nanocomposite surfaces. The aptasensor exhibited a detection limit of 23 fM for thrombin, good sensitivity, and high specificity [123].

Based on the hemin/GQs that can simultaneously act as NADH oxidase and peroxidase DNAzyme, different hemin/GQ DNAzyme electrochemical biosensors for the detection of thrombin were reported [124–127].

A pseudo triple-enzyme cascade electrocatalytic electrochemical biosensor for the determination of thrombin used the amplification of alcohol dehydrogenase-Pt-Pd nanowires bionanocomposite and hemin/GQ structure that simultaneously acted as NADH oxidase and peroxidase DNAzyme [127]. The alcohol dehydrogenase immobilized on the Pt-Pd nanowires catalyzed the ethanol present in the electrolyte into acetaldehyde, accompanied by NAD+ being converted to NADH. Then the hemin/GQ acted first as NADH oxidase, converting the produced NADH to NAD+, and then the hemin/GQ complex acting as peroxidase DNAzyme catalyzed the reduction of the produced H2O2.

Another strategy used porous platinum nanotubes labelled with glucose dehydrogenase and hemin/GQ complexes acting as both NADH oxidase and peroxidase DNAzyme, which led to a cascade signal amplification and allowed the detection limit of thrombin down to 0.15 pM level [128].

The Pebrine disease related Nosema bombycis spore wall protein was detected at a hemin/GQ DNAzyme electrochemical biosensor [129], using the amplification of hemin/GQ DNAzyme functionalized with Pt-Pd nanowires, that again acted as both NADH oxidase and peroxidase, the sensor exhibiting good linear range and detection limit.

Several reports on hemin/GQ DNAzyme electrochemical biosensors used the electrocatalytic properties of the DNAzyme to detect the activities of enzymes and their substrates. In fact, the first hemin/GQ DNAzyme electrochemical biosensor study followed the glucose oxidase (GOx) activity, by attaching the GOx to the gold electrode surface through a nucleic acid sequence able to form GQs in the presence of hemin, Scheme 7 [84]. The GOx mediated the glucose oxidation to gluconic acid and H2O2, and the H2O2 detection by its electrocatalyzed reduction by the DNAzyme. The concentration of H2O2, generated upon the interaction of the modified electrode with glucose, was proportional to the concentration of glucose, and the system could quantitatively determine the glucose substrate. In another report, a multiple current amplification strategy for glucose detection, based on hollow PtCo nanochains functionalized by DNAzyme and GOx, as well as Fc-labelled secondary TBA sequences on hollow PtCo nanochains functionalized by DNAzyme and GOx, proved to be able to distinguish the target protein from interfering molecules.

Mammalian Argonaute 2 (Ago2) protein is the key player of RNA-induced silencing complexes, regulating gene function through RNA interference. A hemin/GQ DNAzyme electrochemical biosensor for the detection of Ago2 protein and study of its RNA endonuclease activity was developed [131]. A hairpin ODN structure, which contained a GQ domain that recognized the hemin and a domain that specifically recognized the target mRNA sequences complex with Ago2, was immobilized onto gold electrode surface. In the presence of Ago2, the hairpin was cleaved into two pieces and the hemin region become free and formed a stable hemin/GQ complex in the presence of K+ ions. The results showed that Ago2 catalyzed the cleavage of target RNA in the absence of any biological partners or ATP and maintained the catalytic activity, in a wide range of pH and temperature.

The adenosine deaminase (ADA) activity was detected at a hemin/GQ DNAzyme electrochemical biosensor [132], which employed an ODN sequence with three functional
domains, an adenosine aptamer domain, a GQ domain, and a linker domain, immobilized on a gold electrode. In the presence of adenosine, the adenosine aptamer formed a close-packed tight structure with the adenosine. However, upon addition of ADA in the test solution, adenosine was converted into inosine due to the catalytic reaction, and the release of inosine made the adenosine aptamer region flexible again. The ADA inhibition was also studied in the presence of erythro-9-(2-hydroxy-3-nonyl) adenine hydrochloride inhibitor.

A DNA-based electrochemical method for the detection of alkaline phosphatase (AP) activity has been developed that used a DNA polymerase terminal deoxynucleotidyl transferase (TdT) and hemin/GQ DNAzyme nanowires acting as both NADH oxidase and peroxidase [133]. A 3'-phosphorylated double-stranded DNA probe was immobilized on a gold nanoparticles modified GC electrode. In the presence of AP, the 3'-phosphoryl end of the DNA probe becomes dephosphorylated, and the TdT catalyzed the DNA probe extension with a poly(T) sequence. Then, a G-rich DNA strand was hybridized with the poly(T) sequence of the DNA probe, which then formed a hemin/GQ DNAzyme in the presence of hemin. In the presence of NADH, the hemin/GQ DNAzyme oxidized NADH to NADH, accompanied by the formation of H$_2$O$_2$, which was further catalyzed by the DNAzyme. The developed biosensor presented good sensitivity, selectivity, reproducibility, and stability, showing promising practical applications in AP activity assay.

(b) Cu$^{2+}$/GQ DNAzyme. Besides hemin, it was demonstrated that the human telomeric DNA assembled with Cu$^{2+}$ ions can present DNAzyme activity, being able to catalyze the Friedel-Crafts reaction in water with excellent enantioselectivity [134, 135]. Based on a Cu$^{2+}$/GQ DNAzyme, an electrochemical method for pyrophosphatase (PPase) activity detection was developed [136]. In the absence of PPase, Cu$^{2+}$ coordinated with pyrophosphate (PPI) to form a Cu$^{2+}$-PPI compound. In the presence of PPase, the PPase catalyzed the hydrolysis of PPI into inorganic phosphate and produced free Cu$^{2+}$, which self-assembled to the G-rich DNA on the screen printed gold electrode surface and formed a Cu$^{2+}$/GQ DNAzyme. Using 3,3',5,5'-tetramethylbenzidine as a redox mediator, the Cu$^{2+}$/GQ DNAzyme catalyzed the reduction of H$_2$O$_2$ to generate a quantitative chronoamperometric signal. This method was additionally applied to screen the sodium fluoride inhibitor for PPase [136].

3.4. Cancer Cells. With the increase demand on the development of new strategy for cancer early detection, GQ electrochemical biosensors have the potential to be important tools for cancer cell detection in early cancer diagnosis. A hemin/GQ DNAzyme electrochemical biosensor for the detection of human liver hepatocellular carcinoma cells (HepG2) was proposed [137], based on the thiolated TLS11a aptamer attached to a gold electrode specific recognition of the target HepG2 cells. Hemin/GQ modified gold nanoparticles were also used for current amplification. After the electrochemical detection, the activation potential of −0.9 to −1.7 V was used to regenerate the gold electrode surface, the gold electrode showing good reusability.

In another approach, the HepG2 cells were detected at sandwich-type hemin/GQ DNAzyme electrochemical biosensor [138], with thiolated TLS11a aptamers attached to the gold nanoparticles modified GC electrode surface. The hemin-GQs were immobilized on Au-Pd core-shell nanoparticle-modified magnetic Fe3O4/MnO2 beads (Fe3O4/MnO2/Au-Pd). The hemin/GQ DNAzymes catalyzed the oxidation of hydroquinone with H$_2$O$_2$, amplifying the electrochemical current and improving the detection sensitivity.

4. Conclusions

This review discussed the recent advances on the electrochemical characterization, design, and applications of G-quadruplex electrochemical biosensors in the evaluation of metal ions, G-quadruplex ligands, and other small organic molecules, proteins, and cells. The incubation time and cations concentration dependence in controlling the G-quadruplex folding, stability, and formation of complex quadruplex-based nanostructures at the surface of carbon electrodes were discussed.

The different G-quadruplex electrochemical biosensors design strategies: the detection, via redox labels, of the G-rich DNA probe folding into a G-quadruplex structure after binding the analyte, the use of G-quadruplex aptamers, as recognition elements to capture the analyte, and the use of hemin/G-quadruplex DNAzymes, for electrochemical current amplification, were revisited.

G-quadruplex aptamer and/or DNAzymes-based biosensing strategies hold great promise for future applications in various fields, ranging from medical diagnostics and treatment to environmental monitoring and food safety. When compared with immunoassay-based biosensors, aptamer and/or DNAzymes-based electrochemical biosensors are particularly promising for the detection of small molecular
targets, since it is difficult to produce highly specific antibodies for small molecules.

Understanding the G-quadruplex folding and stability at the solid-liquid interface of the electrochemical transducers is one of the fundamental challenges of the G-quadruplex electrochemical biosensors development and applications. The electroanalytical characterization using combined electrochemical and surface characterization techniques, of the G-quadruplexes redox behaviour and adsorption process, is crucial and emerges as an important and necessary step for the development of new, more sensitive, G-quadruplex electrochemical biosensors. More powerful signal amplification strategies are also estimated to be developed, in order to overcome the small concentrations of the molecular targets in real medical and environment samples, which may limit the biosensor sensitivity.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

Financial support from Fundação para a Ciência e Tecnologia (FCT), Grant SFRH/BPD/92726/2013 (A.-M. Chiorcea-Paquim), Project UID/EMS/00285/2013 (cofinanced by the European Community Fund FEDER), FEDER funds through the program COMPETE, Programa Operacional Factores de Competitividade, and Innovec'EAU (SOE1/P1/F0173) are gratefully acknowledged.

**References**


[87] G. Liang and X. Liu, “G-quadruplex based impedimetric 2-hydroxyfluorene biosensor using hemin as a peroxidase enzyme.”


Research Article

Putative HIV and SIV G-Quadruplex Sequences in Coding and Noncoding Regions Can Form G-Quadruplexes

Petra Krafčíková, Erika Demkovičová, Andrea Halaganová, and Viktor Víglaský

Department of Biochemistry, Institute of Chemistry, Faculty of Sciences, P. J. Safarik University, 04001 Kosice, Slovakia

Correspondence should be addressed to Viktor Víglaský; viktor.viglasky@upjs.sk

Received 30 July 2017; Revised 26 September 2017; Accepted 18 October 2017; Published 31 December 2017

Academic Editor: Shozeb Haider

Copyright © 2017 Petra Krafčíková et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The HIV virus is one of the most studied viruses in the world. This is especially true in terms of gene sequencing, and to date more than 9 thousand genomic sequences of HIV isolates have been sequenced and analyzed. In this study, a series of DNA sequences, which have the potential to form G-quadruplex structures, is analyzed. Several such sequences were found in various coding and noncoding virus domains, including the U3 LTR, tat, rev, env, and vpx regions. Interestingly, a homologous sequence to the already well-known HIV integrase aptamer was identified in the minus-strand. The sequences derived from original isolates were analyzed using standard spectral and electrophoretic methods. In addition, a recently developed methodology is applied which uses induced circular dichroism spectral profiles of G-quadruplex-ligand (Thiazole Orange) complexes to determine if G-rich sequences can adopt G-quadruplex structure. Targeting the G-quadruplexes or peptide domains corresponding to the G-rich coding sequence in HIV offers researchers attractive therapeutic targets which would be of particular use in the development of novel antiviral therapies. The analysis of G-rich regions can provide researchers with a path to find specific targets which could be of interest for specific types of virus.

1. Introduction

The human immunodeficiency virus (HIV) is an RNA retrovirus in the Retroviridae family which causes HIV infection and over time can lead to acquired immunodeficiency syndrome (AIDS). The virus belongs to the single-stranded positive-sense RNA Lentivirus genus and is formed from two molecules of genomic RNA that are converted into double-stranded DNA by the viral reverse transcriptase. The resulting viral DNA is then inserted into the cellular DNA by the HIV integrase. Once integrated, transcription from the proviral promoter at the 5′ long terminal repeat (LTR) generates mRNAs which code viral proteins and genomic RNA. The integrated provirus may become latent and the infected host cell can remain undetected by the immune system. To date, more than nine thousand HIV-1, SIV, and HIV-2 subtypes have been completely or partially sequenced. The genomic HIV contains regions rich in G-residues which show a marked tendency to adopt G-quadruplex structures, and a number of studies published in recent years have revealed the biological significance of G-quadruplexes in HIVs [1–7]. G-quadruplexes have long been seen as a highly promising target for the development of new anticancer therapies [8, 9], but recent work has also suggested the possibility of adopting a similar strategy for the development of antiviral therapies. Formation of G-quadruplex is usually linked to moderate DNA transcription [10, 11]. Research devoted to G-quadruplexes has so far been limited to the field of viral genomes, despite the advantages of their small size and often naturally occurring double-stranded circular (episomal) form. One important study found that the G-quadruplex Epstein–Barr virus could disrupt the interaction of EBV nuclear antigen 1 with RNA. The linking regions of EBNA1 LR1 and LR2 were revealed to have a strong preference for G-quadruplex RNA and it was revealed that G-quadruplex RNA-interacting drugs block the functions of EBNA1 that are critical for viral DNA replication and episome maintenance [12, 13].

Recently it was confirmed that consensus sequences forming stable G-quadruplex structures are responsible for RNA replication and inhibition of protein translation of hepatitis C virus [14].
Our recently published results have also highlighted the significance of some G-rich regions in regulating areas with the ability to form stable G-quadruplexes in papilloma viruses [15].

G-quadruplex structures also seem to be critical for HIV-1 infectivity and could represent novel targets for antiviral drug development. For example, it is known that mutations disrupting G-quadruplex formation can enhance HIV promoter activity in cells and that treatment with G-quadruplex ligands decreases promoter activity and displays antiviral effects [6]. The U3 region contains a G-rich sequence ∼80 nucleotides upstream from the transcription-starting site (TSS) and close to the TATA box. This sequence overlaps three SP1 binding sites which play a crucial role in the initiation of transcription [1, 4]. Recent research has confirmed the interaction between the SP1 protein and a fragment of the HIV-1 promoter sequence folded into a G4 [16], and the effect of point mutations which disrupt the G-quadruplexes formed in the promoter has been analyzed [6].

Retroviral RNAs are now known to dimerize via G-rich regions in the cytoplasm of infected cells allowing two copies of the genome which is encapsidated in the newly produced virion [17]. The mechanisms which drive RNA dimerization play a role during the strand transfers which may be partially responsible for the viral variability through the production of recombinant molecules [18, 19].

G-rich sequences can form bimolecular G-quadruplex structures in the gag region of the HIV-1 genome, in close proximity to the dimer initiation site (DIS) [16, 20–22]. It has recently been shown that the recombination in the U3 domain is cation-dependent and is significantly lower in the presence of lithium ions, ions which are known to destabilize G-quadruplexes [23].

G-quadruplexes derived from the sequence of the negative regulatory factor (Nef) of HIV-1 were recently analyzed in vitro [5]. Nef G-quadruplexes repress the Nef expression and this finally results in a decrease in viral replication. Thus, targeting the G-quadruplexes located in the Nef coding sequence could lead to further attractive therapeutic opportunities.

Therefore, the main goal of this study is to scrutinize HIV provirus genomes in an attempt to find G-rich regions which may be prone to forming G-quadruplex motifs. Several tools and strategies are available to predict G-quadruplex propensity from some sequences, but there are disadvantages and limitations associated with each algorithm [24–27]. Within the last ten years, it has become generally accepted that stable G-quadruplexes are mainly formed in G-rich regions consisting of four G-runs that contain two or more continuous guanosine residues (G\(_2\_k\) \(n \geq 3\)) and one dinucleotide island GG; see more details in Material and Methods. In fact, the existence of stable G-quadruplexes containing only 3 continuous G-runs has recently been confirmed, and this sequential motif can form so-called bulged G-quadruplexes [31]. Many such sequences have been found in various regions of HIV genomes, and those which are analyzed in this study are summarized in Figure 1. The formation and structure of G-quadruplexes of each oligonucleotide were verified using UV and CD spectroscopy and electrophoretic separation in the presence of either sodium or potassium ions. In order to exclude the false confirmation of G-quadruplex formation on the base of the CD spectra profiles alone, CD melting curves were also determined because the stability of all known G-quadruplexes is significantly higher in the presence of potassium than in the presence of sodium ions.

### 2. Material and Methods

All chemicals and reagents were obtained from commercial sources. DNA oligomers were obtained from Metabion, Germany (Figure 1). PAGE purified DNA was dissolved in double distilled water prior to use. Thiazole Orange was purchased form Sigma-Aldrich (cat. number 390062). Single-strand concentrations were determined precisely by measuring absorbance (∼260 nm) at 95°C using molar extinction coefficients [15]. DNA concentration was determined using UV measurements carried out on a Jasco J-810 spectropolarimeter (Easton, MD, USA). Cells with optical path lengths of 10 mm were used, and the temperature of the cell holder was controlled with an external circulating water bath.


The search criteria for G-quadruplex forming sequences were restricted to sequences which possessed three continuous G-runs containing at least three neighboring Gs and one G-run containing only two neighboring Gs. We aimed to identify sequences with 1–4 nucleotides occurring between two continuous G-runs and with fewer than 9 nucleobases between G-runs in total; thus, the total required number of Gs was set at a minimum of 12. Initially, the reading frame of DNA was adjusted to 20 nucleotides (Figure 2).

The sequences fulfilling these criteria were considered as putative G-quadruplex forming sequences. If an additional G-run was located in close proximity (i.e., less than 3 nucleotides) to the putative sequence, it was also judged to be suitable for inclusion. In principle, we applied criteria similar to those utilized by QGRS mapper and the more comprehensive mining tool QuadBase2, a program which predicts G-quadruplex forming G-rich sequences (QGRS) in nucleotide sequences [27, 32]. The scores of G-quadruplex putative sequences found in HIV genomes were also analyzed using G4Hunter strategy (Figure 1) [25]. It is important to note that QGRS mapper and G4Hunter strategies can miss many sequences which were found by our access.
The randomly selected complete genomic sequences of 20 different HIV-1, 5 HIV-2, and 5 SIV viruses were analyzed, thereby identifying the sequences listed in Figure 1. The name of the oligonucleotide was derived from the type of virus and the first letters of ID; the abbreviations H1-, H2-, and S- represent HIV-1, HIV-2, and SIV, respectively. Bioinformatic analysis was then performed by analyzing the occurrence of each sequence in other genomes using the Basic Local Alignment Search Tool (BLAST), a tool which can identify regions of local similarity between sequences [33]. The sequences summarized in Figure 1 were compared with the complete and partial genome sequences of all known HIVs available in the NCBI Gene database. Oligonucleotides numbers 1, 2, 4, 5, 6, 9, and 10 at the 3’-termini contain an additional proximal G₃-run, and our criteria allow this to be included in the final sequence. However, the sequences marked by rectangles in such cases can represent the supposed loops of “standard” G-quadruplexes consisting of four Gₙ-runs (n ≥ 3). However, the sequences of oligonucleotides used in experimental measurements are too short for relevant sequence alignment, as this approach is primarily used to examine sequences located in long terminal repeat regions (LTRs). Therefore, the wider regions were also used for bioinformatics analyses by adjusting the frame which is restricted by the NF-κ-B and TATA boxes for sequences located in LTRs. A similar methodology was also used for the translation products of the sequences located in coding regions.

### Table 1: DNA Oligonucleotides Used in This Study Originating from HIVs and SIVs

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>nt</th>
<th>e^5</th>
<th>Sequence 5' → 3'</th>
<th>Subtype Region</th>
<th>Genbank ID</th>
<th>G4H Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>H2-U22</td>
<td>27</td>
<td>312.0</td>
<td>TGGGGAGGGCACTGGGCGGT</td>
<td>HIV-2 Vpx</td>
<td>U22047.1</td>
<td>1.85</td>
</tr>
<tr>
<td>(2)</td>
<td>H2-M15</td>
<td>26</td>
<td>306.0</td>
<td>TGGGGAGGGCACTGGGCGGT</td>
<td>HIV-1 LTR (Sp1)</td>
<td>M15390.1^5</td>
<td>1.88</td>
</tr>
<tr>
<td>(3)</td>
<td>H2-U38</td>
<td>23</td>
<td>267.0</td>
<td>TGGGGAGGGCACTGGGCGGT</td>
<td>HIV-2 Vpx</td>
<td>U38293.1^5</td>
<td>1.78</td>
</tr>
<tr>
<td>(4)</td>
<td>H1-JX</td>
<td>26</td>
<td>302.0</td>
<td>TGGGAAGTTAGGGGCGGTTC</td>
<td>HIV-1 LTR (Sp1)</td>
<td>U22047.1</td>
<td>1.55</td>
</tr>
<tr>
<td>(5)</td>
<td>H1-JX1</td>
<td>26</td>
<td>300.0</td>
<td>TGGGAAGTTAGGGGCGGTTC</td>
<td>HIV-2 LTR (Sp1)</td>
<td>J04498.1^1</td>
<td>1.76</td>
</tr>
<tr>
<td>(6)</td>
<td>H1-L20</td>
<td>27</td>
<td>317.0</td>
<td>TGGGAAGTTAGGGGCGGTTC</td>
<td>HIV-2 LTR (Sp1)</td>
<td>M15390.1^5</td>
<td>1.63</td>
</tr>
<tr>
<td>(7)</td>
<td>H2-U38B</td>
<td>20</td>
<td>250.0</td>
<td>AGGGACTGAGGCTGGGGTAG</td>
<td>HIV2 LTR (Sp1)</td>
<td>M30931.1^5</td>
<td>1.72</td>
</tr>
<tr>
<td>(8)</td>
<td>H2-J0</td>
<td>21</td>
<td>254.0</td>
<td>GGGCGGCCATGGGCAAGAC</td>
<td>HIV1 LTR (Sp1)</td>
<td>JX12863.1^5</td>
<td>2.27</td>
</tr>
<tr>
<td>(9)</td>
<td>H2-M15B</td>
<td>30</td>
<td>362.0</td>
<td>AGGGACTGAGGCTGGGGTAG</td>
<td>HIV1 LTR (Sp1)</td>
<td>JX12863.1^5</td>
<td>2.05</td>
</tr>
<tr>
<td>(10)</td>
<td>S-M30</td>
<td>25</td>
<td>280.0</td>
<td>GGGCGCGTATGGGCTAGGA</td>
<td>HIV-1, SIV LTR (Sp1)</td>
<td>U22047.1</td>
<td>1.71</td>
</tr>
<tr>
<td>(11)</td>
<td>S-JX</td>
<td>22</td>
<td>242.0</td>
<td>CGGGCGCGTATGGGCTAGGA</td>
<td>HIV-1, SIV LTR (Sp1)</td>
<td>J04498.1^1</td>
<td>1.50</td>
</tr>
<tr>
<td>(12)</td>
<td>H1-JN</td>
<td>21</td>
<td>234.0</td>
<td>TGGGCGAGCTGGGAGTGGC</td>
<td>HIV1 LTR (Sp1)</td>
<td>M30931.1^5</td>
<td>1.72</td>
</tr>
<tr>
<td>(13)</td>
<td>H1-K02</td>
<td>22</td>
<td>242</td>
<td>GTGGCTGGGCGGGACTGGGG</td>
<td>HIV1 LTR (Sp1)</td>
<td>J04498.1^1</td>
<td>1.50</td>
</tr>
<tr>
<td>(14)</td>
<td>H1-M27</td>
<td>22</td>
<td>239.0</td>
<td>GTGGCTGGGCGGGACTGGGG</td>
<td>HIV1 LTR (Sp1)</td>
<td>M30931.1^5</td>
<td>1.72</td>
</tr>
<tr>
<td>(15)</td>
<td>H1-K03</td>
<td>18</td>
<td>202.0</td>
<td>CGGGTGCTGGGAGTGGG</td>
<td>HIV1, SIV LTR (Sp1)</td>
<td>J04498.1^1</td>
<td>1.50</td>
</tr>
<tr>
<td>(16)</td>
<td>H1-JN-A</td>
<td>22</td>
<td>257.0</td>
<td>AGGGACTGAGGCTGGGGTAG</td>
<td>HIV1 LTR (Sp1)</td>
<td>JX12863.1^5</td>
<td>2.05</td>
</tr>
</tbody>
</table>

*Occurrence in other organisms containing at least one point-mutation. ^Millimolar extinction coefficient at 264 nm (mM⁻¹ cm⁻¹). ^G4 Hunter coefficient [25]. ^Number of hits: •, 1–10; ••, 11–99; •••, 100–1000; ••••, 1000–3500; Sp1 consensus sequence 5’-(G/T)GGGCGG(G/A)(G/A)(C/T)-3’ is underlined; G-runs are highlighted by red color. The rectangular selection represents the supposed loops of G-quadruplexes consisting of four Gₙ-runs (n ≥ 3).
**Figure 2**: Strategy and searching criteria of putative G-quadruplex sequences. Randomly selected genomic sequences of HIV-1, HIV-2, and SIV from NCBI Gene database were analyzed (panel (a)). The criteria used to determine putative G-quadruplex sequences are listed in panel (b). $\sum S_i$ represents the total number of nucleotide residues between G-runs ($G_3$ and $G_2$), and this value is restricted at the interval $\langle 3 – 9\rangle$. Specific criteria were applied in cases in which G-runs consisted of more than four Gs (c); for instance, $G_7$ is formed of either two $G_3$ islands interrupted with one $G$ (yellow square) or one $G_3$ and $G_2$ connected with two “yellow” Gs. The occurrence of putative sequences was researched by BLAST in other viruses listed in the Gene database. Positive and negative examples of the search procedure are shown in panels (d) and (e).
sodium chloride. pH was adjusted by TRIS to a final value of 7.0.

DNA titration was performed with increasing concentrations of Thiazole Orange (TO). TO was solubilized in DMSO to reach a final concentration of stock solution of 10 mM. The concentrations of DNA and TO in a 1 mm quartz cell were 30 μM and 0–200 μM, respectively, and the increment of TO was ~67 μM. Each sample was mixed vigorously for 3 min following the addition of TO; CD/UV spectra were measured immediately.

2.3. CD Melting Curves. CD melting profiles were collected at ~295 and ~265 nm as a function of temperature, using a procedure which has been published previously [33]. The temperature ranged from 0 to 100°C, and the heating rate was 0.25°C per minute. The melting temperature ($T_m$) was defined as the temperature of the midtransition point.

2.4. Thermal Difference Spectra. The conditions and parameters used in the examination of the thermal difference spectra were identical to those used in the CD spectroscopy assay. The spectra analysis performed in this study has been described in an earlier publication [34].

2.5. Electrophoresis. Native polyacrylamide gel electrophoresis (PAGE) was performed in a temperature controlled vertical electrophoretic apparatus (Z375039-1EA, Sigma-Aldrich, San Francisco, CA). Gel concentration was 12% (19 : 1 monomer to bis ratio, Applichem, Darmstadt). Approximately two micrograms of DNA was loaded onto 14 × 16 × 0.1 cm gels. Prior to loading, each DNA sample was heated to 95°C for 5 min in an appropriate buffer and cooled to room temperature. Electrophoreses were performed at 20°C for 4 hours at 120 V (~8 V · cm⁻¹). DNA oligomers were visualized with Stains-All immediately after electrophoresis, and the electrophoretic record was photographed on a white pad with a Nikon D3100 camera. The gel was also later stained by silver staining procedure in order to improve the sensitivity of the DNA visualization [15, 35].

3. Results and Discussion

3.1. Bioinformatic Analysis. Although many different HIV sequence comparisons have been performed to date, this study offers an alternative means of identifying putative G-quadruplex forming sequences. The search criteria were not restricted to the LTR regions of the proviral HIV genome, but were instead applied to the entire proviral HIV genome. In this overview, the occurrence of 16 selected oligonucleotide sequences within more than nine thousand previously sequenced HIV/SIV genomes was examined in an attempt to identify some general relationship between them. The sequence structure consisting of three G-runs containing at least three neighboring Gs and one G-run containing two Gs was found in HIV-1, HIV-2, and SIV provirus DNAs. These sequences and their sources are summarized in Figure 1. Some sequences were also found in other organisms, but some of these are obligatory, located only in an appropriate HIV genome (Figure 3).

3.1.1. The Vpx Region. The sequence H2-U22 was found only in one HIV-2 isolate in the terminal part of the vpx region (ID:
U22047.1), but its derivatives containing 1–3 point mutations were found in an additional 9 HIV-2 isolates. H2-U38 is a truncated version of H2-U22, and this sequence was found again in the same region in an additional five HIV-2 genomic sequences (ID: M30502.1, U38293.1, M31133.1, U22047.1, and KU168289.1). The first 20 nucleotides of H2-M15 are identical to those of the H2-U22 and H2-U38 sequences, and this oligomeric sequence occurred very rarely in HIV's, being found in only 2 isolates of HIV-2 in vpx region (ID: X05291.1, M15390.1) and two derivatives containing 1-2 mutations. Interestingly, both sequence derivatives can also be found in other organisms. Considering the extreme rarity of these three sequences in HIV-2 genomes, their significance and biological role are questionable.

3.1.2. Env/Rev Region. The G-rich region located in the env gene of HIV-1 is also a promising potential source of G-quadruplex formations. The env and rev coding sequences overlap, but their reading frames are different. This region contains the HI-JN-A sequence which occurs in only 11 HIV-1 isolates, but a derivative in which the central guanosine is substituted for adenosine (AGGGACT-GAGACTGGGTTGGGA) occurs in more than 1000 HIV-1 isolates. Interestingly, our results confirm that the formation of G-quadruplex is not affected and this sequence adopts preferentially the dimer form (not shown). There is some analogy here between the abasic site in the second G-run and the G for A substitution which has been the subject of recent studies by two different groups [36,37]. The abasic site and the G for A substitution decrease the thermodynamic stability of such derivative sequences. Therefore, this substitution might not be sufficient to prevent such sequences from forming G-quadruplex motifs. It should be possible to verify the influence of a G for A substitution in the second G-run on a series of oligonucleotides. These results lead us to form the hypothesis that the formation of G-quadruplexes, as with the formation of other secondary motifs, could lead to a pausing effect on the DNA replication, transcription, or translation of the env and rev regions [38–40].

A number of different research projects have attempted to identify conserved structural motifs in highly variable viruses which can be used as specific targets for the development of efficient antiviral therapies. Interestingly, HI-JN-A sequence encodes the oligopeptides Gly–Leu–Arg–Leu–Gly–Trp–Glu and Gly–Thr–Glu–Ala/Thr–Gly–Val–Gly which are integral parts of Env and Rev proteins, respectively. These oligopeptide motifs are highly abundant in HIV-1 proteins and were found in more than 1140 coding sequences of Env and Rev.

3.1.3. The Minus-Strand. All of the sequences described thus far are found in the plus-DNA/RNA strand. However, the sequence 5'-ACCCACCTCCCACACCCCCG-3' is typically located in the plus-strand of HIV-1 at the beginning of the second exon of tat/rev and env genes. This motif is complementary to the sequence 5'-d(CGCGGTGGAGTGGG-3') in the minus-strand and, interestingly, is very similar to the well-known HIV-93del aptamer d(GGGGT-GGGAGGGG), which forms very stable interlocked dimeric G-quadruplex [41]; the two sequences differ in two bases. Additionally, 98 known HIV-1 isolates differ only in one extra thymine compound between the first two G-runs, for example, isolate KU168259.1 in Gene bank. The HI-K03 sequence is located in N160 various HIV-1 isolates.

Although the homology of HI-K03 and aptamer sequences is undoubtedly interesting, we are unable to offer a convincing explanation for the phenomenon. Nevertheless, this is the first reported case of a natural coding sequence being homological to an aptamer which was originally developed against the protein produced by the same organism. Is this merely a coincidence or is it an exception? However, if the sequence was located in the coding strand, it would be possible to elucidate an explanation or a convincing theory about the biological role of the sequence.

3.1.4. LTR Regions. In recent years, a wide range of research and publications has focused on the study of G-rich sequences in LTR. In principle, the results of our research into U3 LTR sequences fully corroborate the earlier findings of other authors [2–4, 6, 7]. The HI-L20, HI-JX, and HI-JX1 sequences are very similar; indeed, HI-JX and HI-JX1 differ only in one central nucleobase. These sequences were found in HIV-1 genomes in LTR Spl region. Their occurrences are 8 and 9 hits for HI-JX and HI-JXI, respectively. Their derivatives were found in more than 30 various HIV-1 isolates. HI-L20 was found in only 5 isolates, although variants containing 1-2 point mutations were located in an additional 28 isolates. The formula implying all possible variants of HIV isolates which overlap HI-L20, HI-JX, and HI-JXI sequences is

\[ 5'\text{-GGGGAGGRAYRDKGGYGGDSDGGGA-3'} \]

where the following nomenclature of wobbles in DNA nucleobases is used: \(R = (A \text{ or G}), B = (C \text{ or G or T}), Y = (C \text{ or T}), D = (A \text{ or G or T}), S = (G \text{ or C}), H = (A \text{ or C or T}), W = (A \text{ or T}), V = (A \text{ or C or G}), K = (G \text{ or T}), M = (A \text{ or C}), \text{ and N can be any base.} \]

The HI-JN sequence was found in a significantly higher number of HIV-1 variants, 3181 hits in HIV-1 genomic sequences and only one in the SIV isolate. This sequence was not found to occur in HIV-2 and other organisms. Recent studies have identified and described the structure of HIV-1 sequence LTR-IV: d(CTGGGCCGGACTGGGAGTGGT) and their derivatives [42], and the underlined nucleotides of this sequence are homological to HI-JN sequence. The parallel G-quadruplex containing the bulge was confirmed by NMR analysis (PDB ID: 2N4Y).

HI-K02 and HI-M27 sequences are highly homological. These sequences partially overlap with HI-JN; their occurrence in HIV-1 genome was again found to be very high and is identified in more than 1700 various isolates in the NCBI database. The first guanosine is highly conservative in both sequences. This guanosine could be essential for the formation of G-tetrads and may also contribute to the stability of G-quadruplexes exhibiting bulge features. The large size of the statistical set would suggest the likelihood of higher numbers of nucleobase variations, and this was confirmed by the BLAST analysis of the region occurring between NF-κ-B and TATA boxes of KJ849802.1, in which
H1-JN was found to occur in 1061 various isolates. All possible variants express the formula

$$5'\text{-N-RDDVNHGDVSHGRRRHNNRGADKSVB-3'}.$$  (2)

The number of sequenced HIV-2 genomes was markedly smaller in comparison to that of HIV-1 and this inevitably resulted in a smaller statistical set. The sequence H2-U38B is a truncated version of H2-M15B; with the exception of the terminal 5'-adenosine of H2-J0, the two sequences are a perfect overlap. Their occurrence in different isolates was not as frequent as H1-JN, possibly due to the smaller number of sequenced HIV-2 genomes; all the identified sequences occurred in the LTR of HIV-2 and not in HIV-1. The truncated H2-U38 and H2-J0 derivatives were not only limited to humans, but also found in many other organisms such as Ovis canadensis, Macaca fascicularis, Mus musculus, Rattus norvegicus, Mesocricetus auratus, and Fundulus heteroclitus and pigs and plants such as Solanum lycopersicum and Orzya punctate. 145 different isolates of HIV-2 were found in the wider region determined with the NF-κ-B and TATA boxes. The formula including all possible variants is as follows:

$$5'\text{-DRRGWRRNNRYTRGRRGDYRKGKRRGG-3'}. $$  (3)

The total number of all known HIV-1 sequences is much higher than the number of sequenced SIV genomes, but although the S-JX sequence originally found in SIV occurred prevalently in HIV-1 isolates, this was not the case for S-M30 sequence. Both S-JX and S-M30 sequences were also located in the LTR Sp1 region. S-M30 was found only in one SIV isolate, but its derivatives were found in an additional 6 SIV genomes and in other organisms including Pseudomonas, elephant endotheliotropic herpesvirus, Enterobacter cloacae, Streptomyces leuvenhoekii, and Leptosphaeria maculans lepidii.

The sequence alignment of the region containing the sequences used in this study, located in LTR of HIV-1 and SIV, is summarized in Figure 4. The alignments of the wider region determined by NF-κ-B and TATA boxes are shown in the Supporting Information (Table S1).

The formula including all possible variants is as follows:

$$5'\text{-NDRGGRHGGGRRYYDRGGAGTGG-3'}. $$  (4)

The models describing the formation of four G-quadruplexes formed in the LTR region have been described in a recent study [1, 6, 7]. The structure of one of these, an antiparallel G-quadruplex structure composed of only two tetrads, was also confirmed using NMR [4]. These four topologies are mutually exclusive because the G-runs associated with the formation of a single G-quadruplex are overlaid with others, thereby preventing the formation of any of the three alternative conformations. This sequence of four G-runs or more arranged in tandem is not uncommon in viral genomes, and similar regions consisting of 5-7 G-runs have also been found in human papillomaviruses [15]. Thus, the equilibrium between these forms may play a role in regulating promoter activity in viruses.

Targeting G-quadruplexes including the possible variations located in the LTR coding sequence of HIVs can therefore offer an attractive therapeutic opportunity for the development of highly efficient inhibitors of processes depending on the secondary motifs in this regulating region.

### 3.2. CD Measurements

Another aim of this study was to confirm the ability of the studied oligonucleotides to form stable G-quadruplexes. In order to ascertain this, a series of experiments using circular dichroism analysis was performed (Figure 5). DNA oligonucleotides were analyzed in a buffer supplemented with 50 mM of potassium and sodium ions. All the studied oligonucleotides showed signatures which are typical for the formation of G-quadruplex structures in the presence of 50 mM KCl; positive CD peaks were recorded at ~265 nm and/or 295 nm. The comparison of relative molar intensity of CD peaks with human telomeric repeats sequence at 265 or 295 nm shows that the sequence which folds into G-quadruplex can consist of three tetrads [34]. All CD spectra and CD melting profiles obtained in the presence of sodium and potassium are summarized in Supporting Figure S1.

### 3.3. Electrophoretic Analysis

Information about the molecular and the presence of multimeric conformers of G-quadruplexes can be obtained by examining samples using electrophoretic separation [43]. Electrophoreses were performed in the presence of 50 mM KCl and NaCl at 20°C.
Figure 5: CD spectra of HIV oligomers in modified 25 mM mBR buffer (pH 7.0) in the presence of 50 mM KCl. The corresponding UV, TDS, and CD melting curves obtained at 265 and 293 nm are shown in Supporting Figures S1 and S2.
Figure 6: Molecular standard S, the mix of d(AC)$_9$, d(AC)$_{14}$, and d(AC)$_{18}$, was used. Electrophoretic separation was performed in a 14% polyacrylamide gel at 10°C in 25 mM Britton-Robinson buffer (pH 7.0) and 50 mM KCl at 8°C in (a) and 50 mM NaCl in (b). Prior to being used, the DNA sample was heated in the same buffer for 5 min at ~98°C and slowly cooled to room temperature within 30 min.

Table 1: Melting temperatures and molecularities of the studied DNA oligonucleotides in the presence of 50 mM sodium and potassium ions.

<table>
<thead>
<tr>
<th>Oligo.</th>
<th>Fold</th>
<th>$T_m$ [°C]</th>
<th>Potassium</th>
<th>Fold</th>
<th>$T_m$ [°C]</th>
<th>Sodium</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) H2-U38</td>
<td>M, D</td>
<td>56.8$^p$</td>
<td>M, D</td>
<td>52.4$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) H2-U22</td>
<td>M</td>
<td>56.6$^p$</td>
<td>M $\gg$ D</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) H2-M15</td>
<td>M</td>
<td>57.1$^p$</td>
<td>M, D</td>
<td>46.6$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) H1-JX</td>
<td>M</td>
<td>57.5$^p$</td>
<td>M $&gt; D$</td>
<td>30.3$^p$, 52.5$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) H1-JX1</td>
<td>M</td>
<td>61.6$^p$</td>
<td>M $&gt; D$</td>
<td>35.8$^p$, 51.8$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) H1-L20</td>
<td>M</td>
<td>58.2$^p$</td>
<td>M</td>
<td>36.4$^a$, 50.4$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) H2-U38B</td>
<td>M $&lt; D$</td>
<td>59.5$^p$, 96.5$^p$</td>
<td>M</td>
<td>55.3$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) H2-J0</td>
<td>M, H $\ll$ D</td>
<td>51.8$^p$, 870$^p$</td>
<td>M, H $\ll$ D</td>
<td>50.8$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) H2-M15B</td>
<td>M</td>
<td>46.5$^p$</td>
<td>M$^{46} \gg D$</td>
<td>44.2$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) S-M30</td>
<td>M</td>
<td>67.8$^p$</td>
<td>M</td>
<td>49.6$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) S-JX</td>
<td>M, D$^f$</td>
<td>54.2$^a$, 79.5$^p$</td>
<td>M, D</td>
<td>35.0$^a$, 39.2$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(12) H1-JN</td>
<td>M</td>
<td>48.3$^a$, 46.3$^p$</td>
<td>M $&gt; D$</td>
<td>38.7$^a$, 43.7$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13) H1-K02</td>
<td>M</td>
<td>48.4$^a$</td>
<td>M $&gt; D$</td>
<td>32.1$^p$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14) H1-M27</td>
<td>M</td>
<td>39.6$^a$</td>
<td>M</td>
<td>ND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15) H1-K03</td>
<td>D, H$^f$</td>
<td>60.3$^a$, &lt;98$^p$</td>
<td>M$^a$, D</td>
<td>39.7$^a$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16) H1-JN-A</td>
<td>M</td>
<td>58.6$^p$</td>
<td>D $\gg M^{46}$</td>
<td>37.2$^p$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$Smear represents flexible conformers; $^b$the mobility can correspond to unfolded structure; M: monomer; D: dimer; H: high ordered multimer; $^p$melting curve was obtained at ~264 nm; $^f$melting curve was obtained at ~293 nm. The error in $T_m$ determination is ±3°C.

(Figure 6). DNA oligomers d(AC)$_9$, d(AC)$_{14}$, and d(AC)$_{18}$ were used as molecular standards. Under identical conditions to those used in spectral measurements, most sequences were found to form exclusively intramolecular structures in the presence of potassium ions (Figure 6(a)). These forms move faster than the molecular standard representing the unfolded ssDNA. However, in the presence of sodium ions, smaller populations of dimeric forms are typically identified (Figure 6(b)). Dimeric structures in the presence of potassium were also observed for H2-U38, H2-U38B, H2-J0, and H1-K03. The results are summarized in Table 1.

Interestingly, the fastest band of H1-K03 represents a dimeric conformer, while the slowest and middle bands correspond to high-ordered structures. Preliminary NMR data indicates that H1-K03 forms an interlocked G-quadruplex structure which is analogous to HIV integrase aptamer [41]. The smearing of the middle band can be attributed to the presence of conformers with similar thermodynamic stabilities which possibly form one or more states during electrophoretic separation. The mobility of the dimer form of H2-M15B in the presence of sodium corresponds to the mobility of the unstructured molecular standards; therefore it is not possible to determine whether this form is a folded dimer or an unfolded dimer based on the electrophoretic record alone. However, the CD melting curve is clearly defined; this is possible only when the unknown
structure transitions from one state to another; therefore it is possible to conclude that this band should represent a dimer structure (see Supporting Information, Figure S1).

It is important to note here, however, that not each band in a certain electrophoretic column necessarily represents a G-quadruplex structure.

3.4. TDS Analysis. The profiles of thermal difference spectra (TDS) of G-quadruplexes are highly specific, and therefore this analysis was also performed on the studied oligonucleotides, although, as our previous studies have noted, this technique is not wholly reliable and may provide erroneous results [34, 44]. The results are summarized in Supporting Figure S2 and in Table 1. The TDS results for H2-J0, H2-U38B, and H2-M15B were somewhat ambiguous but indicated the possible formation of G-quadruplexes in the presence of both sodium and potassium ions. Interestingly, their TDS show a maximum close to 260 mM in the presence of sodium, while the local minimum at 295 nm is not obvious in the presence of both sodium and potassium. Other oligonucleotides show a local minimum at 295 nm in TDS; for more information, see Supporting Figure S2.

More reliable results can, however, be obtained from melting curve analysis. The methodology of this technique is based on the fact that all known G-quadruplexes are more stable in the presence of potassium than in the presence of sodium ions [45]. The results of the melting curve analysis for our set of oligonucleotides revealed that the same sequences which offered interesting results in TDS assay, H2-J0, H2-U38B, and H2-M15B, also shared intriguing features in their melting curve profiles. Each of these sequences showed identical melting temperatures regardless of whether sodium or potassium ions were present in the used buffer. This raises the unanswered question of what type of secondary structure is formed within this set of oligonucleotides. A recent study described a novel tetrahelical structural motif which is distinct from the typical G-quadruplex structure, but this form has similar spectral properties as we observed in our experiments for these three oligonucleotides [46]. This study analyzed G3AGCG repeats found in the regulatory region of the PLEKHG3 gene and found that the VK sequences d(G3AGCGA)nG3AGCG, where n = 1, 3, are capable of forming tetrahelical DNA stabilized with unusual noncanonical base-pairings: G-G and G-A. In addition, the sequences HPV25 and HPV25/2 found in the E4 gene of human papillomavirus type 25 also share many similar signatures [15]. These sequences prefer to adopt another fold as a G-quadruplex. Based on these findings, we decided to perform a sequence alignment of HPV25, HPV25/1, H2-M15B, H2-J0, and H2-38B to assess the homology of the three sequences. The results show a relatively high homology among the sequences, which indicates that these three HIV-2 sequences (H2-J0, H2-U38B, and H2-M15B) could form structures very close to those of the HPV25 sequence (see Supporting Table S2). The melting temperatures of VK and HPV25 sequences were not dependent on the type of metal cation [15, 46]. Extrapolating from this, it is possible to suggest that the structure may form a hairpin-like structure containing unusual G-G and G-A basepairs as was confirmed in VK. In principle, the tetrahelical VK structure is a special case of crossed hairpin structure.

It has been shown that polyethylene glycol 200 (PEG200) stabilizes G-quadruplexes in the presence of potassium and destabilizes double helical motifs [45]. However, PEG-200 significantly increases the melting temperatures of the H2-M15B, H2-J0, H2-38B, VK, and HPV25 sequences in the presence of potassium; see Supporting Figure S5. In order to solve this Gordian knot and determine whether the H2-M15B and H2-38B sequences adopt G-quadruplex or other structural motifs, we decided to use another experimental method.

3.5. CD Titration Analysis. Recently, a newly developed experimental methodology using the ligand Thiazole Orange (TO) for the identification of G-quadruplex forming sequences has been applied [47]. TO has a strong binding affinity to triplexes and G-quadruplexes, an affinity which is significantly higher to them than to other structural motifs [48, 49]. Although TO is optically inactive, TO-quadruplex complexes are chiral and display a unique profile of the induced CD (ICD) spectrum in the visible region [47]. This methodology offers valuable results in a wide range of conditions, but it is most sensitive in solutions without the presence of metal cations. Similarly, it can also be applied with slightly reduced sensitivity in solutions containing Na+ or low concentrations of K+. Nevertheless, we performed the titration experiments in the presence of both 50 mM KCl and 50 mM NaCl because these concentrations of salts are more biologically relevant. The representative results of the titration analysis of H1-JN with TO in the presence of 50 mM NaCl and KCl are shown in Figure 7.

The ICD results display the expected positive signals at ~495 and ~510 nm and the negative signals at ~475 nm. These signatures are characteristic for TO-quadruplexes complexes [47]. The titration of all sequences is shown in Supporting Figures S3 and S4.

As expected, each oligonucleotide was also found to have formed G-quadruplexes under the given conditions. Signals corresponding to those of G-quadruplex structures were also clearly detected in the UV region. In case of antiparallel G-quadruplexes, the signals at 295 and 265 nm were seen to decrease and increase, respectively, by increasing the concentration of TO, phenomena which are indicative of the conversion from antiparallel to parallel folding.

The titration analysis of the HPV25 sequence shows that ICD was mirrored; the positive peaks become negative and vice versa. It is therefore possible to assume that the binding mode of TO with this sequence must be different than those typical for G-quadruplex motifs (Figure 8). The same effect was also observed for VK sequence. These results indicate that H2-M15B, H2-J0, and H2-U38B are also able to form G-quadruplex structures.

4. Conclusion

Our bioinformatic study of HIV genomes partially corresponds with the analyses recently published by many authors in that it too focuses primarily on G-rich regions located in U3 LTRs. However, this study reveals that G-quadruplexes
Figure 7: The representative CD titration spectra of ~27 µM HI-JN with TO. 0, 2.5, 5, and 7.5 molar equivalents of TO represent black, green, blue, and red lines, respectively. Each sample was measured in modified 25 mM Britton-Robinson containing (a) 50 mM KCl and (b) 50 mM KCl.

Figure 8: CD titration spectra of 27 µM HPV25/1 (d(GGGAGCGGGAC-TGGGACCGGGACCG-GG)) with TO. 0, 2.5, 5, and 7.5 molar equivalents of TO are represented by black, green, brown, and red lines, respectively. Each sample was measured in a modified 25 mM mBR buffer containing 50 mM NaCl.

Several unanswered questions require deeper analysis to determine the features that provide specific G-quadruplex motifs with the ability to function as structural elements.

In this study, we used only the cost-effective methods to confirm that some oligonucleotides form G-quadruplex motifs. However, we again demonstrate that ICD signal of TO-quadruplex complex offers valuable additional information, allowing distinguishing whether an unknown sequence has ability to adopt G-quadruplex structure.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Slovak Research and Development Agency under Contract no. APVV-0280-11, European Cooperation in Science and Technology (COST CM1406), Slovak Grant Agency 1/0131/16, and internal university grants (VVGS-PF-2017-251 and VVGS-2016-2596). The authors thank G. Cowper for critical reading and correction of the manuscript.

Supplementary Materials

Table S1: the sequence alignment of the regions determined by NF-κ-B and TATA boxes located in LTRs containing the sequences used in this study. Table S2: sequence alignments of HPV25, HPV25/1, H2-M15B, H2-J0, and H2-38B. Figure S1: the condition of CD spectra and melting curve measurements are the same as in Figure 3. Measurements in the presence of 50 mM NaCl and 50 mM KCl are represented as blue and black lines, respectively. The melting curves were obtained at 265 nm (solid lines) and 294 nm (dotted lines). The relative CD intensity is compared with the reference HTR oligonucleotide d(G3T2A)3G3. The value of 1 is obtained when the molar CD signal at 295 nm is the same as the CD signal of HTR at the same concentration [Tóthová et al. Biochemistry 53, 2014, 7013-27]. Figure S2: thermal difference spectra in the presence of sodium (blue lines) and potassium (black lines). Figure S3: the CD titration spectra of 27 µM DNA sample with TO; 0, 2.5, 5, and 7.5 molar equivalents of TO are represented by black, green, blue,
and red lines, respectively. Each sample was measured in a modified 25 mM mBR buffer containing 50 mM NaCl. Figure S4: CD titration spectra of 27 μM DNA sample with TO. 0, 2.5, 5, and 7.5 molar equivalents of TO are represented by black, green, blue, and red lines, respectively. Each sample was measured in a modified 25 mM mBR buffer containing 50 mM KCl. Figure S5: CD spectra and melting curves of H2-M15B, VK, and HPV25/1 sequences in 50 mM KCl and the presence (magenta) and absence of 50% v/w PEG200 (black lines). (Supplementary Materials)

References


Research Article

Telomeric G-Quadruplexes: From Human to Tetrahymena Repeats

Erika Demkovičová,1 Ľuboš Bauer,1 Petra Krafšiková,1 Katarína Tlučková,1,2 Petra Tóthova,1 Andrea Halaganová,1 Eva Valušová,3 and Viktor Viglaský1

1Department of Biochemistry, Institute of Chemistry, Faculty of Sciences, P. J. Šafárik University, 04001 Košice, Slovakia
2Department of Biological Sciences/RNA Institute, University at Albany, SUNY, Albany, NY 12222, USA
3Institute of Experimental Physics, Slovak Academy of Sciences, Watsonova 47, 040 01 Košice, Slovakia

Correspondence should be addressed to Viktor Viglaský; viktor.viglasky@upjs.sk

Received 30 July 2017; Revised 11 November 2017; Accepted 5 December 2017; Published 28 December 2017

Academic Editor: Shozeb Haider

Copyright © 2017 Erika Demkovičová et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The human telomeric and protozoal telomeric sequences differ only in one purine base in their repeats; TTAGGG in telomeric sequences; and TTGGGG in protozoal sequences. In this study, the relationship between G-quadruplexes formed from these repeats and their derivatives is analyzed and compared. The human telomeric DNA sequence G(TA)3 and related sequences in which each adenine base has been systematically replaced by a guanine were investigated; the result is Tetrahymena repeats. The substitution does not affect the formation of G-quadruplexes but may cause differences in topology. The results also show that the stability of the substituted derivatives increased in sequences with greater number of substitutions. In addition, most of the sequences containing imperfections in repeats which were analyzed in this study also occur in human and Tetrahymena genomes. Generally, the presence of G-quadruplex structures in any organism is a source of limitations during the life cycle. Therefore, a fuller understanding of the influence of base substitution on the structural variability of G-quadruplexes would be of considerable scientific value.

1. Introduction

G-rich DNA sequences can form intra- and intermolecular G-quadruplexes based on the association of one or more DNA strands. The nucleotides which intervene between G-runs form loops of folded G-quadruplex structures which can adopt a variety of different topological forms [1,2]. When the guanine tracts are oriented in the same direction, the double-chain reversal (propeller) loops link two adjacent parallel strands to form a parallel structure [3]. When the guanine tracts are oriented in opposite directions, the edgewise or diagonal loops link two antiparallel strands to form an antiparallel G-quadruplex [4]. In antiparallel hybrid or so-called (3 + 1) structures, a single strand is oriented in a different direction from the others [5–7]. A novel (3 + 1) type fold which has recently been described by Marušič et al. exhibits a conformation in which all three loop types occur in one conformation: edgewise, diagonal, and double-chain reversal loops [8]. In addition, intermolecular multimeric G-quadruplexes can be formed by the association of two or more strands [9].

These structures underline the high degree of G-quadruplex structural polymorphism, a phenomenon which is dependent on many different factors: the length and sequence of nucleic acid, and environmental conditions present during the folding reaction such as the buffer, pH, stabilizing cation, temperature, and the presence of agents causing dehydration [10–14]. G-rich sequences with the propensity to form G-quadruplex structures can be located in many regions of human genomic DNA, especially in several biologically important regions including the end of linear eukaryotic telomeres [15,16]. However putative G-rich sequences are not randomly distributed within a genome; such sequences predominantly occur in protooncogene regions (which promote cell proliferation) and are depleted in tumour suppressor genes (which maintain genomic stability) [17]. It is very
unlikely that these putative sequences can form in vivo and direct evidence of their existence in living cells is still a topic of discussion [18–20]. Undoubtedly, the most extensively studied G-quadruplex forming sequences are those located at the 3’-ends of human telomeres. Telomeric sequences and specialized nucleoprotein complexes which cap the ends of linear chromosomes are essential for chromosomal stability and genomic integrity [21–23]. Mammalian telomeres consist of tandem repeats of G-rich sequences, d(TTAGGG), and genomic imperfections were also found in the human and Tetrahymena genomes; see Table 1 and Supporting Materials.

The structure of human telomeric DNA in crowded solutions has also been investigated by many authors [11], but this structure is likely to be a result of dehydration rather than molecular crowding [12, 32, 33]. The great variety of structures identified to date can also be attributed to the presence of flanking nucleotides outside the core sequence G₃(T₂A₂G₃)₃ and the concentration of ions and to the use of different experimental methods and conditions [1, 34].

A series of systematic studies concerning the sequence derivatives of human telomeric repeats were carried out by Vorlícková et al. [35–38], and these earlier studies focused on the substitution of guanine for adenine, the introduction of abasic sites, 8-oxoadenine replacing adenine, and the substitution of 5-hydroxymethyluracil for thymine in telomeric repeats were analyzed [39–41]. However, in this study, an opposite strategy is applied, the substitution of adenine for guanine (see Figure 1). The main aim is to achieve the total conversion of four human repeats to Tetrahymena repeats which retain the ability to form intramolecular G-quadruplex. Interestingly, G-rich repetitions containing imperfections were also found in the human and Tetrahymena genome; see Table 1 and Supporting Materials.

In this study, we examine the structures formed by the Tetrahymena telomeric sequence, dG₃(T₂A₂G₃)₃, which differs from the human sequence, dG₃(T₂A₂G₃)₃, in the substitution of adenine for guanine. The HTR to THR conversion by substitutions of adenine for guanine is shown in Figure 1: The HTR to THR conversion by substitutions of adenine for guanine.

### Table 1: DNA oligodeoxynucleotides used in this study and their occurrence in human and Tetrahymena genomes.

<table>
<thead>
<tr>
<th>Name</th>
<th>ε₄ (mM⁻¹ cm⁻¹)</th>
<th>Sequence (5' → 3')</th>
<th>Occurrence in genome (ID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTR</td>
<td>225.4</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>Very high -</td>
</tr>
<tr>
<td>HTR₁</td>
<td>222.1</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NG 054915.1 M116271.1</td>
</tr>
<tr>
<td>HTR₂</td>
<td>222.1</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NG 029533.1 -</td>
</tr>
<tr>
<td>HTR₃</td>
<td>222.1</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NT 187653.1 -</td>
</tr>
<tr>
<td>HTR₁₂</td>
<td>218.8</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NW 00357049.1 M116271</td>
</tr>
<tr>
<td>HTR₁₃</td>
<td>218.8</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NC 018926.2 M116271</td>
</tr>
<tr>
<td>HTR₂₃</td>
<td>218.8</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NC 018914.2 M116271</td>
</tr>
<tr>
<td>HTR₀₁</td>
<td>230.0</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>- M116271</td>
</tr>
<tr>
<td>HTR₀₃</td>
<td>230.0</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NC 018930.2 M116271</td>
</tr>
<tr>
<td>THR</td>
<td>215.4</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NC 018921.2 AH001112.2</td>
</tr>
<tr>
<td>THR₁₃</td>
<td>213.4</td>
<td>GGGTTAGGTTAGGGTTAGGG</td>
<td>NG 034020.1 Very high</td>
</tr>
<tr>
<td>G4C₂</td>
<td>204.4</td>
<td>GGGCCGCGCGCGCGCGGG</td>
<td>ND ND</td>
</tr>
<tr>
<td>(AC)₉</td>
<td>171.9</td>
<td>(AC)₉</td>
<td>ND ND</td>
</tr>
<tr>
<td>(AC)₁₈</td>
<td>342.9</td>
<td>(AC)₁₈</td>
<td>ND ND</td>
</tr>
</tbody>
</table>

*a* Milimolar extinction coefficient at 257 nm. *b* Base modifications are underlined. ND: not determined.
from the human sequence by a single G-for-A replacement in each repeat [42]. Since Gs are essential for the formation of G-quadruplexes, we have systematically substituted each of the three adenines for guanines in the TTA loops of the G-quadruplex-forming sequence G₃(T₂A₃G₃), thereby increasing the number of guanines by up to three guanines per oligonucleotide. Circular dichroism spectroscopy (CD) and polycrystalamide gel electrophoresis (PAGE) were used to observe the effect of base substitution (s) on the formation, thermal stability, and conformation of G-quadruplexes. The measurements were performed in the presence of both Na⁺ and K⁺ ions and with concentrations of either PEG-200 or acetonitrile at 0, 15, 30, and 50 wt% at different temperatures. In addition, the formation of G-quadruplex structures was verified and confirmed using Thiazole Orange (TO). TO is an excellent DNA fluorescent probe for DNA structural forms because of its high fluorescence quantum yield [43]. This ligand stabilizes the G-quadruplex structure and can also induce topological changes [44, 45]. The G-quadruplex-TO complex offers a characteristic profile of induced-circular dichroism spectrum in buffers containing sodium cations [44].

2. Materials and Methods

All experiments were carried out in a modified Britton-Robinson buffer (mRB), 25 mM phosphoric acid, 25 mM boric acid, 25 mM acetic acid, and supplemented by 50 mM of KCl or NaCl, PEG-200 (polyethylene glycol with an average molecular weight of 200) and acetonitrile (Fisher Slovakia); pH was adjusted by Tris to a final value of 7.0. Oligonucleotides with sequences shown in Table 1 were purchased from Metabion international AG. The lyophilized DNA samples were dissolved and diluted in suitable buffers containing sodium cations [44].

2.1. CD Spectroscopy. CD and UV-vis spectra were measured using a Jasco model J-810 spectropolarimeter (Easton, MD, USA). The temperature of the cell holder was regulated by a PTC-423L temperature controller. Scans were performed over a range of 220–600 nm in a reaction volume of 300 µl in a cuvette with a path length of 0.1 cm and an instrument scanning speed of 100 nm/min, 1 nm pitch, and 1 nm bandwidth, with a response time of 2 s. CD data represents three averaged scans taken at a temperature range of 0–100°C. All DNA samples were dissolved and diluted in suitable buffers containing appropriate concentrations of ions and dehydrating agent. The amount of DNA oligomers used in the experiments was kept close to 25 µM of DNA strand concentration. The samples were heated at 95°C for 5 minutes then allowed to cool down to the initial temperature before each measurement. CD spectra are expressed as the difference in the molar absorption of the right-handed and left-handed circularly polarized light (Δe) in units of M⁻¹·cm⁻¹. The molarity was related to DNA oligomers. A buffer baseline spectrum was obtained using the same cuvette and subtracted from the sample spectra. The thermal stability of different quadruplexes was measured by recording the CD ellipticity at 295 and 265 nm as a function of temperature [14, 46]. The temperature ranged from 0 to 100°C, and the heating rate was 0.25°C/min. The melting temperature (Tₘ) was estimated from the peak value of the first derivative of the fitted curve. DNA titration was performed with increasing concentrations of TO. TO was solubilized in DMSO to reach a final concentration of stock solution of 10 mM. The concentration of DNA and TO in 1 mm quartz cell was 30 µM and 0–200 µM, respectively, and the increment of TO was ~67 µM. Each sample was mixed vigorously for 3 min following the addition of TO; CD/UV spectra were measured immediately.

2.2. Electrophoresis. Samples consisting of 0.3 µl of 1 mM stock solutions were separated using nondenaturing PAGE in a temperature-controlled electrophoretic apparatus (Z375039-1EA; Sigma-Aldrich, San Francisco, CA) on 15% acrylamide (19:1 acrylamide/bisacrylamide) gels. DNA was loaded onto 13 × 16 × 0.1 cm gels. Electrophoresis was run at 10°C for 4 hours at 125 V (~8 V-cm⁻¹). Each gel was stained with StainsAll (Sigma-Aldrich). The gel was also stained using the silver staining procedure in order to improve the sensitivity of the DNA visualization [44].

2.3. Fluorescence Spectroscopy. The fluorescence spectra were acquired with a Varian Cary Eclipse Fluorescence Spectrophotometer at 22 ± 1°C which was equipped with a temperature-controlled circulator. A quartz cuvette with a 3 mm path length was used in all of the experiments. In the fluorescence measurements, the excitation and emission slits were 5 nm and the scan speed was 240 nm/min. 66 µM of TO was titrated with DNA (3.3, 6.6, and 13.2 µM) in a mRB buffer in both the presence and absence of monovalent metal cations. The molar ratios between DNA and ligand were 1:20, 1:10, and 1:5. The excitation wavelength was adjusted to 452 nm.

3. Results and Discussion

3.1. Sequence Design and CD Spectra. The sequence derived from human telomeric sequence d(G₃(T₂A₃G₃)) and substituted derivatives under different conditions are studied. The DNA sequences and the abbreviations used in this study are summarized in Table 1. Points 1, 2, and 3 indicate the positions of the base substitution in the first, second, and third loops of the HTR sequence, respectively. Point 0 indicates a flanking guanine at the 5’ end of the oligonucleotide, Figure 1. In the DNA oligonucleotides derived from HTR, the guanine (G)-for-adenine (A) in the TTA loop was substituted with the expectation that the modified sequences would retain the ability to form G-quadruplexes spontaneously, albeit with different topologies than those found in HTR sequences. The HTR derivatives were analyzed in the presence of both 50 mM NaCl and KCl, Figure 2. The first group represents oligonucleotides containing only single point mutations at different positions; HTR₁, HTR₂, and HTR₃ (black lines in Figure 2). The second group represents oligonucleotides...
Figure 2: CD spectra of oligonucleotides used in this study in a 25 mM modified Britton-Robinson buffer (pH 7.0) in the presence of 50 mM KCl (a–c) and 50 mM NaCl (d–f). The HTR and THR spectra are shown in red and magenta, respectively. Each DNA sample was annealed at 95°C for 5 min and then allowed to cool for ~1 h to the initial temperature at which the sample was kept at the beginning of the measurement [14].
containing two point mutations (spectra indicated with blue in Figure 2). The first two loops were modified in HTR\textsubscript{1,2}, the first and last loops were changed in HTR\textsubscript{1,3} and the second and third loops were modified in HTR\textsubscript{2,3}. Oligonucleotides HTR\textsubscript{0,1,3}, HTR\textsubscript{0,2,3}, and HTR\textsubscript{1,2,3} contained three G-for-A substitutions (spectra in green). The spectrum and melting temperatures of the HTR\textsubscript{0,1,2} sequence are very similar to those of the HTR\textsubscript{0,2,3} sequence (not shown in this study), while the HTR\textsubscript{0,1,2,3} sequence is equivalent to the THR sequence.

The substituted sequences were also compared with the unmodified HTR and THR sequences. In general terms, each of the guanine residues in any G-run could be involved in the formation of G-tetrads. In the case of the formation of three-layered G-tetrad quadruplexes, loop lengths were found to vary when the base substitution was introduced into the HTR sequence; loops could consist of three or four nucleotides depending on the location and number of substitutions. However, we cannot exclude the possibility of the formation of four-layered G-quadruplexes for sequences containing three substitutions, but it is important to note that such structures would have to consist of at least one heteronucleotide-tetrad in which adenine is also present. To date, the 3D structure of full-length THR sequences in presence of potassium has not been determined; the only facet of the structure which is known is the tetrameric G-quadruplex structure formed from four shorter sequences d(GGGGCC) (PDB: 201D and 230D) adopted a structure with two edgewise and one central diagonal loops [4, 51, 52].

This structure consists of four G-tetrads and cannot be stated as representing the real structure of a full-length oligonucleotide. Nevertheless, the 3D structure of THR has been ascertained only in the presence of sodium (PDB: 186D) [48]. This structure consists of three stacked G-tetrads, two edgewise loops, and one double-chain-reversal loop. Despite the fact that the sequences of THR and HTR differ at only one of the six nucleotides, their 3D topologies are quite different because HTR in sodium adopts a three-G-tetrad structure consisting of two edgewise loops and one central diagonal loop (PDB: 143D) [4]. However, the HTR sequence can also adopt a stable basket-type conformation in the presence of potassium consisting of only two G-tetrad layers (PDB: 2KF8) [31].

Several naturally occurring HTR sequences have been identified to date. Forms I (PDB: 2HY9) and 2 (2JPZ) consist of three G-tetrads, but the order of loops differs; HTR forms one double-chain-reversal and two edgewise loops in both forms [49, 50]. There is some similarity with THR G-quadruplexes which form in solution in the presence of sodium [48]. Form 3 is represented by a parallel G-quadruplex with three double-chain-reversal loops (PDB: 1KF1) [3]. Recently, Lim et al. have also confirmed the structure of a 27-nt HTR derivative in the presence of sodium which differs significantly from those mentioned above (PDB: 2MB) [1]. Although both known HTR structures solved in sodium possess the same relative strand orientations, they differ in the hydrogen-bond directionalities and in the loop arrangement. The 2MB structure again consists of two edgewise and one double-chain-reversal loops.

The sequence derived from the telomere of Oxytricha d[G\textsubscript{4}(T\textsubscript{4}G\textsubscript{4})\textsubscript{3}](PDB: 201D and 230D) adopts a structure with similar types of loops to those found in HTR in sodium; two edgewise and one central diagonal loops [4, 51, 52]. However, the Oxytricha sequence forms a four-layered G-tetrad quadruplex. At the time of writing, the solution structure of Oxytricha sequence d[G\textsubscript{4}(T\textsubscript{4}G\textsubscript{4})\textsubscript{3}] in K\textsuperscript{+} containing solution had yet to be determined. The main reason for this could be the fact that this sequence and THR in the presence of potassium can adopt different topological forms which coexist in solution; additional bands are observed during electrophoretic separation [14]. Interestingly, the four-layered G-quadruplexes are very stable, exhibiting particularly high melting temperatures in the presence of potassium [14]. Recently, the structure of d(GGGGCC)\textsubscript{3} in the presence of potassium has also been determined; the sequence contains cytosines instead of thymine residues and one 8-bromodeoxyguanosine (PDB: 2N2D) [53]. The G-quadruplex structure adopted by this sequence could be closely related to that of THR in potassium. This antiparallel structure is composed of four G-quartets which are connected by three edgewise C-C loops. CD spectra results show many signatures in common with the THR sequence. One of the cytosines in every loop is stacked upon the G-quartet; an arrangement which results is a very compact and stable structure. Similarly, the melting temperature of the structure is higher than 90°C.

It is generally accepted that CD spectroscopy is a very useful and cost-efficient method for offering a first glance at the architecture of folded G-quadruplexes. CD spectra of G-quadruplexes can be used to indicate whether the DNA has folded into a parallel or antiparallel conformation [36, 54].

Although there are up to 25 generic folding topologies of G-quadruplexes, it is possible to classify the structures into three groups based on the sequence of glycosidic bond angles adopted by guanosines of the G-quadruplex [55]. Group I consists of parallel G-quadruplexes with strands oriented in the same direction and with guanosines of the same glycosidic bond angles. Parallel G-quadruplexes (Group I) share the same characteristics irrespective of whether they contain three or four loops: an intense positive maximum at ~265 nm and minimum at ~240 nm. Groups II and III consist of antiparallel G-quadruplexes; Group II can be characterized by guanosines of glycosidic binding angles in orientations such as anti-anti and syn-syn and also syn-anti and anti-syn, while Group III consists of stacked guanosines of distinct glycosidic bonding angles. Antiparallel G-quadruplexes show a positive band at ~295 nm. Positive and negative CD signals at ~265 nm at ~240 nm, respectively, are characteristic for Group II, while Group III shows reverse peaks. In contrast, the CD spectra of high ordered G-quadruplex architecture of Group III forms exhibit negative and positive signals at 240 nm and ~265 nm, respectively [55].

CD profiles corresponding to distinct G-quadruplex conformational states are determined empirically; therefore, the interpretation of CD spectra of unknown putative G-quadruplex sequences can be ambiguous. A number of other factors can also cause a degree of uncertainty over the evaluation of CD spectra, including, for example, the presence of mixed populations of various conformers and/or the presence of multimeric conformations in solution [9, 14, 44, 46].
CD measurements clearly show that the G-for-A substitutions had a considerable impact on the spectral profile of each sequence. The presence of the G-quadruplex scaffold formed from the unmodified HTR sequence is characterized by a positive peak at ~295 nm with two shoulders at around ~270 and ~250 nm in the presence of potassium (Figure 2(a), red line). According to CD spectra these signatures are characteristic for Group II antiparallel G-quadruplexes. This spectrum is indicative of the formation of a two-layered basket-type structure [31, 55]. The HTR sequence adopts a clear antiparallel G-quadruplex conformation of Group II type in the presence of sodium (Figure 2(d), red line). The structure is characterized by a large positive maximum at ~295 nm, a smaller one near ~245 nm, and a negative CD peak at ~265 nm. Previous studies have reported that these sequences form an intramolecular, basket-type antiparallel G-quadruplex [4]. Every sequence shows a clear peak at ~295 nm which is characteristic of an antiparallel G-quadruplex topology. The first set of oligomers with a single substitution per oligonucleotide in the presence of potassium shows two separated peaks at ~295 and ~265 nm; the signal is dominant at 295 nm (spectra shown in black). However, THR and HTR derivatives containing one or more G-for-A substitutions in the presence of potassium show an increase of the peak at ~265 nm (Figures 2(b) and 2(c)). This indicates the coexistence of more than one topological structure, that is, both parallel and antiparallel configurations; see also the electrophoretic results in Figure 7. The structural polymorphism was seen to increase with increasing numbers of Gs in the DNA sequence. The CD signal at ~265 nm (spectrum shown in green) was predominant for oligonucleotides containing three substitutions (Figure 2(c)).

In the presence of sodium, only the HTR0 sequence with a substitution in the second loop exhibited a CD spectrum identical to that of HTR, although even this correspondence displayed lower amplitudes (spectra in dotted black in Figure 2(d)). HTR1 and HTR3 sequences with substitutions in the first and third loop, respectively, also displayed a positive maximum at 295 nm, but the negative peaks at 265 nm were shallower and slightly shifted towards longer wavelengths in comparison to the results of the unmodified sequence. Despite these differences, they are nonetheless likely to form G-quadruplexes of Group III. Only the CD spectra of the THR sequence shows signatures of Group II types.

Samples in the second group exhibited a positive maximum at ~295 nm with a slight shift to lower wavelengths in the case of HTR1,2 and HTR1,3 (Group III, spectra shown in blue in Figure 2(b)). These two sequences displayed a lack of a negative peak at 265 nm, and the smaller positive peak at around 245 nm was shifted slightly to longer wavelengths (Figure 2(e)). HTR1,3 shows a negative signal at ~245 nm and a positive signal at 265 and 295 nm, results which are indicative of the formation of Group II antiparallel G-quadruplexes.

The CD spectra of HTR0,1,3 are close to those of Group II G-quadruplexes while the CD of HTR0,1,3 and HTR1,2,3 resemble those of Group III G-quadruplexes. All samples with three mutations exhibited a positive maximum at ~295 nm. HTR0,1,3 exhibits negative signals at 235 and 275 nm in the presence of sodium (Figure 2(f)), while HTR0,1,2,3 shows positive signals at 265 and 295 nm.

In general, all the modified sequences in the presence of both Na+ and K+ were seen to differ to some degree from the HTR spectrum and were also found to differ from each other. The varying CD spectral profiles from sample to sample are a result of slight changes in G-quadruplex topology. However, it was not possible to determine either the group or structure of the G-quadruplexes with any degree of certainty on the basis of CD spectral profiles alone due to the coexistence of various topological forms, a finding which was confirmed by the electrophoretic results discussed in Section 3.5.

3.2. CD Spectra in the Presence of PEG-200 and Acetonitrile.

In the presence of K+, the dehydrating agent PEG-200 is known to induce a conformational change of telomeric G-quadruplexes, primarily the transition from an antiparallel structure to a parallel arrangement [2, 9, 11, 13, 14, 56]. Therefore, the influence of PEG-200 and another dehydrating agent, acetonitrile on CD spectral results and the stability of HTR derivatives were also investigated. The representative CD spectra of HTR and THR in the presence of different concentrations of both dehydrating agents (15, 30, and 50 wt%) and 50 mM KCl are shown in Figure 3. Both types of DNAs were found to form G-quadruplex structures with a propeller-like parallel arrangement in the presence of K+. However, when the sequences were studied in the presence of sodium with no potassium present, no structural conversions were observed; this finding remained constant for all of the studied HTR derivatives and THR. Interestingly, at a PEG-200 concentration of 50 wt% the positive peaks at 295 nm were found to disappear and a CD signal at 265 nm was recorded which was ~2-fold higher than without the presence of PEG-200. The same effect was observed for acetonitrile. This is an intrinsic property of any converted G-quadruplex molecule. In a recent study, our group presented a hypothesis which explains this fact; the CD signal depends on the number and orientation of stacked glycosyl bonds [9, 14, 57]. We have also previously shown that PEG-200 causes the dimerization of HTR [9]; therefore electrophoretic analysis of the sequences in the presence of PEG-200 was also performed, Figure 8.

The melting temperature of HTR and the vast majority of G-quadruplex structures are known to increase in the presence of PEG-200. In order to verify this fact, the melting temperatures were determined in the presence of PEG-200 on the basis of CD melting curves. The results are summarized in Table 2 and clearly confirm that PEG-200 increases the melting temperatures of HTR derivatives. In a methodology, which has been used in our previous studies, dual wavelength measurements were performed for cases in which the spectra displayed peaks at both 295 and 265 nm, respectively [9, 11]. A Tm of 63.2°C was obtained in a mRB buffer containing 50 mM KCl, as compared to a value of 50.4°C in a buffer with 50 mM NaCl for HTR at 295 nm. The overall picture which emerges from the thermodynamic data is that the stability of G-quadruplexes of HTR derivatives increases with increased numbers of G-for-A substitutions in both Na+ and
Figure 3: CD spectra of HTR (a and b) and THR (c and d) oligomers in a 25 mM mBR buffer (pH 7.0) in the presence of 50 mM KCl (red lines). The samples contain either PEG 200 or acetonitrile; 15% (v/w) (green lines), 30% (v/w) (light blue lines), and 50% (v/w) (dark blue lines). The increase in the magnitude of CD signals at 265 with increasing concentrations of dehydrating agent is marked by an arrow in both panels. Each DNA sample was annealed at 95°C for 5 min and then allowed to cool overnight at 4°C. The dashed red line represents the spectrum of THR without annealing.

Table 2: Influence of PEG-200 on the melting temperatures of DNA oligomers in the presence of potassium and sodium ions.

<table>
<thead>
<tr>
<th>Oligomer</th>
<th>( T_m ) (°C) in 50 mM KCl + PEG-200</th>
<th>( T_m ) (°C) in 50 mM NaCl + PEG-200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0% 15% 30% 50%</td>
<td>0% 15% 30% 50%</td>
</tr>
<tr>
<td>HTR</td>
<td>63.2/62.9 63.2/ND 78.8/ND ND/89.8 50.4/ND 51.2/ND 52.8/ND 64.0/ND</td>
<td></td>
</tr>
<tr>
<td>HTR(_1)</td>
<td>71.8/73.1 71.4/76.3 64.3/83.1 ND/95 55.7/ND 58.9/ND 61.5/ND 64.4/ND</td>
<td></td>
</tr>
<tr>
<td>HTR(_2)</td>
<td>65.2/65.9 64.2/72.1 ND/79.6 ND/95 51.2/ND 54.3/ND 58.1/ND 62.7/ND</td>
<td></td>
</tr>
<tr>
<td>HTR(_3)</td>
<td>67.8/68.6 69.3/75.3 ND/81.7 ND/95 50.6/ND 54.3/ND 58.7/ND 64.8/71.6</td>
<td></td>
</tr>
<tr>
<td>HTR(_{1,2})</td>
<td>69.5/70.5 ND/76.9 ND/83.2 ND/95 55.9/ND 59.1/ND 62.8/ND 65.3/72.1</td>
<td></td>
</tr>
<tr>
<td>HTR(_{1,3})</td>
<td>72.2/71.5 70.3/79.1 ND/86.4 ND/95 53.7/ND 56.2/59.8 60.1/63.2 ND/73.5</td>
<td></td>
</tr>
<tr>
<td>HTR(_{2,3})</td>
<td>81.3/76.9 ND/79.7 ND/85.2 ND/95 53.3/ND 57.3/ND 59.9/63.0 64.4/71.0</td>
<td></td>
</tr>
<tr>
<td>HTR(_{1,2,3})</td>
<td>73.0/80.0 ND/79.0 ND/87.9 ND/95 52.5/ND 56.3/60.7 60.1/63.3 ND/72.7</td>
<td></td>
</tr>
<tr>
<td>HTR(_{0,1,3})</td>
<td>70.7/72.7 ND/77.6 ND/95 51.8/54.5 56.0/51.1 59.4/60.2 ND/73.8</td>
<td></td>
</tr>
<tr>
<td>HTR(_{0,2,3})</td>
<td>73.9/77.8 ND/83.9 ND/88.6 ND/100 56.4/ND 59.0/ND 61.2/64.0 ND/75.7</td>
<td></td>
</tr>
<tr>
<td>THR</td>
<td>83.1/ND 87.7/ND &gt;95 ND 57.4/ND 59.7/ND ND/66.8 ND</td>
<td></td>
</tr>
</tbody>
</table>
K+ solutions. The lowest $T_m$ value of HTR was recorded in both 50 mM KCl and 50 mM NaCl. The $T_m$ of HTR derivatives was found to be higher in KCl than in NaCl. All of the studied sequences showed a higher $T_m$ value in the presence of both dehydrating agents. The results indicate that PEG-200 stabilizes G-quadruplexes with or without the A-for-G mutation. The proposed melting temperatures summarized in Table 2 clearly demonstrate that both the number of guanine residues in a G-tract and the nature of the stabilizing ion are important determining factors in the thermal stability of G-quadruplexes.

3.3. Titration Measurements. Our group has recently developed a new experimental methodology for the identification of G-quadruplex forming sequences using the cyanine dye Thiazole Orange (TO). TO is an excellent DNA fluorescent probe for various structural motifs due to its high fluorescence quantum yield [58]. This experimental technique can also be used to investigate the hypothesis that HTR derivatives adopt G-quadruplex conformations. TO interacts with various DNA secondary structures, but it has a stronger binding affinity to triplex and G-quadruplex structures than to other structural motifs [43, 45]. Although TO is optically inactive, TO-quadruplex complexes are chiral and display a unique profile of the induced CD (ICD) spectrum in the visible region [44]. Recently we have described the common ICD features shared by many different G-quadruplex structures. The results of TO-quadruplex interaction are the positive peaks at 265 and 295 nm (UV range), and the three peaks in the visible region at $\sim 512$, $\sim 492$, and $\sim 473$ nm in the solution either without the presence of metal cations or in presence of Na+ [44]. TO facilitates the formation of G-quadruplex structures even without the presence of other cations, but the adopted topology induced with TO can vary in comparison with the presence of sodium or potassium in solution; the CD profile in the UV region can be different. A completely different ICD profile of the TO-DNA complex was observed for sequences unable to adopt G-quadruplex structure [44]. However, other G-quadruplex ligands tested in our laboratory were not suitable for this purpose and provided ambiguous results; for example, Thioflavin T, porphyrin derivatives, Hoechst 33342, and Hoechst 33258. This methodology is intended to be used as a supplementary technique because it extends the possibilities of basic spectral methods in terms of distinguishing G-quadruplex structures without the use of more expensive and time-consuming methods. ICD monitoring can be applied in different conditions, but it is the most sensitive in solutions without the presence of metal cations; it can also be applied with slightly reduced sensitivity in solutions containing Na+ or low concentrations of K+ ($<5$ mM). It should be noted here that the interpretation of ICD profile at higher concentration of K+ is by no means unambiguous. Nevertheless, we also performed the titration experiments in the presence of 50 mM KCl because this condition is more biologically relevant.

The results of titration analysis in the presence of 50 mM NaCl are shown in Figure 4. The ICD results display the expected positive signals at $\sim 495$ and $\sim 510$ nm and negative signals at $\sim 475$ nm, signatures which are characteristic for G-quadruplexes. In addition, the G4C2 sequence was analyzed because of its 3D structure in solutions containing potassium. As expected, this oligonucleotide was also found to form G-quadruplexes under the given conditions. Signals corresponding to those of antiparallel G-quadruplex structures were also clearly detected in the UV region. By increasing the concentration of TO, the signals at 295 and 265 nm were seen to decrease and increase, respectively, phenomena which are indicative of a conversion from antiparallel to parallel folding. This effect was also observed under the influence of PEG-200, Figure 3.

As was noted above, titration measurements in 50 mM KCl were also performed. Figure 5. The signals observed in the UV region clearly suggest that G-quadruplex structures were formed in the presence of potassium, but the effect of structural conversion was significantly suppressed. An intensive ICD signal with a maximum of around 500 nm is known to correspond to the formation of complexes between DNA and ligands. However, there was a distinct lack of any of the clear common features which are typically observed for profiles obtained in the presence of sodium. Interestingly, the ICD of the THR sequence was inverted, and therefore we suggest that the binding mode of TO with THR is different from that with HTR derivatives. Another explanation is that THR forms at least two distinct topological conformations in solution and that one of these forms can bind with TO more effectively. As was reported in our previous study, THR forms at least three different structures in the presence of potassium [14]; we therefore decided to verify this hypothesis using electrophoretic separation.

It is important to exclude the potential side effect of using DMSO during TO titration experiments. The stock solution of TO contains DMSO, a polar aprotic solvent which may produce an effect similar to that of PEG-200 [56]. The concentration of DMSO used in our experiments did not exceed 4.5 wt%. In order to eliminate the dehydrating effect DMSO may cause in TO titration experiments, titration analysis was also performed in the presence of DMSO alone. However, no significant effect was observed at concentrations lower than 5 wt% in the absence of Na+, K+, and ions. Nevertheless, the presence of DMSO in solution containing K+ could explain the slight differences in ICD profiles at concentrations of K+ greater than 5 mM.

3.4. Fluorescence Analysis. DNA-TO complexes display a clear but wide absorption at around 500 nm. A single positive peak of TO was observed at 452 nm and this wavelength was used for the excitation of the DNA-TO complex. The fluorescence spectra of the HTR and THR sequences are shown in Figure 6. The measurements were performed in three different environments: (i) a mRB buffer without metal cations, (ii) a mRB buffer supplemented with 50 mM NaCl, and (iii) mRB supplemented with 50 mM KCl. For both oligomers, the fluorescence enhancement achieved the highest yield in the solution without metal cations. The fluorescence yield of HTR was greater than the yield of THR in all three of the tested conditions. The goal of this experiment was to demonstrate that the profile of fluorescence is not greatly dependent on the sequence of DNA oligonucleotide for this ligand. The
fluorescence enhancement of TO can induce the formation of any type of G-quadruplex structure.

3.5. Electrophoresis in the Presence of Na⁺, K⁺, and TO. Nondenaturing polyacrylamide gel electrophoresis (PAGE) is an accessible technique which is used to supplement spectroscopic data when the presence of multiple species of G-quadruplexes cannot readily be identified based on CD spectra alone. The mobility of the DNA sample depends on many different factors, including conformation, charge, and molecular mass. Electrophoretic separation can provide valuable information about the molecularity of G-quadruplexes. Intramolecular G-quadruplexes have a compact structure and thus migrate faster through a cation-containing gel than their linear counterparts, while intermolecular G-quadruplexes migrate more slowly due to their higher molecular weight [9, 14, 44]. Oligomers d(AC)_9, d(AC)_14, and d(AC)_18 were used as standards due to their lack of secondary structures. These standards served as benchmarks in comparing the mobility of different electrophoretic patterns. Since none of the sequences used were longer than 22 nt., the oligonucleotides which were observed to have migrated faster than d(AC)_9 could be identified as having formed intramolecular G-quadruplexes. It is also reasonable to assume that oligonucleotides which moved more slowly or at a similar speed to d(AC)_18 had adopted high-order G-quadruplex structures. Figure 7 shows the electrophoretic records of native 15% polyacrylamide gels illustrating the relative mobilities of the oligomers in the presence of 50 mM NaCl and KCl at 10°C (Figures 7(a) and 7(c)). In addition,
the corresponding electrophoretic results, where the gels and loading buffers contain 2 molar equivalents of TO, are shown in Figures 7(b) and 7(d). In general, some clear trends emerge. Gel electrophoresis performed in the presence of sodium shows that all of the oligonucleotides had moved in one bulk, with single bands migrating faster than d(AC)₄₈ in each column. This effect was also observed when TO was present in the gel. These results indicate that all DNA oligonucleotides form antiparallel intramolecular G-quadruplexes under these conditions. These results agree with the results obtained by CD spectroscopy. It is important to note that intramolecular structures had formed exclusively in the presence of sodium despite the introduction of mutations in HTR sequences increasing the possibility of the formation of different topologies of G-quadruplexes. The electrophoresis did not reveal any significant anomalous mobility of oligomers; sequences with the same length were found to move more or less equally.

In contrast to sodium, the presence of potassium led to the formation of both intra and intermolecular arrangements (Figures 7(b) and 7(d)). In the first group, the HTR₁ quadruplex with one substitution in the first loop exhibited the fastest migrating band in comparison to that of HTR. A single smear band was also observed for the HTR₂ sequence. Smears typically arise when two distinct conformers can be formed; a slow isomerization between the two conformers during the electrophoretic separation is the main source of band smearing. The mobility of the HTR and HTR₃ sequences with substitutions in the third loop is similar. The oligonucleotides containing two substitutions per oligomer displayed high levels of polymorphism. These oligonucleotides form several coexisting conformers because each line contains

Figure 5: CD titration spectra of 27 μM DNA sample with TO. 0, 2.5, 5, and 7.5 molar equivalents of TO represented by black, green, brown, and red lines, respectively. Each sample was measured in a modified 25 mM mBR buffer containing 50 mM KCl.
Figure 6: Fluorescence emission spectra (a.u.) of TO in the presence of HTR (solid lines) and THR (dashed lines) in mRB buffer. The spectra without the presence of metal cations with 50 mM NaCl and 50 mM KCl are represented by black, blue, and red lines, respectively. The molar ratio of DNA:ligand is 1:5. Fluorescence emission of TO is shown in green.

Figure 7: Electrophoretic records of studied DNA oligonucleotides. Electrophoretic gel and buffer contained 50 mM NaCl (a, b) and 50 mM KCl (b, d), gels in (b) and (d) also contained 30 μM TO. 0.4 μL of DNA from 1 mM stock solution was applied to each electrophoretic well (~3 μM). The S-line represents the mobility of the mixture of oligonucleotides: d(AC)₉, d(AC)₁₄, and d(AC)₁₈.
cooled, and then loaded into the electrophoretic wells. 0.4
well DNA from 1 mM stock solution was applied to each electrophoretic

HTR several bands moving at different rates. Interestingly, the presence of PEG-200. Electrophoretic gel and buffer contained 50 mM KCl. DNA samples were heated to 95

Figure 8: Electrophoretic record of DNA oligonucleotides in the presence of PEG-200. Electrophoretic gel and buffer contained 50 mM KCl. DNA samples were heated to 95

and then loaded into the electrophoretic wells. 0.4 µL of DNA from 1 mM stock solution was applied to each electrophoretic well (~3 µM). The loading buffer contained 50 wt% PEG-200.

several bands moving at different rates. Interestingly, the HTR₁₂ and HTR₁₃ sequences displayed two faster well-recognized bands, results which correspond to the formation of intramolecular conformers, and slower bands representing multimeric structures. The addition of TO also caused the fastest conformers to coalesce and the slowest structures to diminish. HTR₂₃ produced a faster intra- and slower intermolecular species (dimer and tetramer). Surprisingly, the oligonucleotides with three substitutions per oligomer were found to be slightly less polymorphic in comparison with the sequences containing two substitutions, displaying only bands with lower magnitudes corresponding to the formation of multiple-molecular G-quadruplexes in the case of the HTR₁₁₁, HTR₂₂₂, and HTR₁₂₃ sequences. TO was found to exert only a limited effect on the multimeric forms of these oligonucleotides.

3.6. Electrophoresis in the Presence of PEG-200. The dependence of HTR dimerization on PEG 200 concentration has been analyzed in previous studies [9]. The formation of both intermolecular dimers and intramolecular monomers was observed in the buffer containing a PEG-200 concentration of 15% wt. The HTR derivatives containing 2 and 3 substitutions were seen to convert readily to slower migrating dimeric structures even at lower concentrations of PEG-200. At a PEG-200 concentration of 50 wt% and 50 mM KCl, the complete structural conversion to a parallel dimeric G-quadruplex was induced, Figures 3 and 8. This effect was not observed in buffers that did not contain potassium [9, 14]. CD measurements at 50 wt% PEG-200 showed no signal at ~295 nm. Based on our previous studies, intermolecular species which migrate more slowly are indicative of the formation of dimers [2, 9, 14]. The 3D structure of HTR containing a flanking sequence in an analogous condition has been determined using NMR [11]. The results show an intramolecular parallel G-quadruplex structure (PDB: 2LD8), but the overhanging nucleotides can cause a steric hindrance for the dimerization of this structure.

4. Conclusion

In this study, we clearly demonstrate that increasing the number of guanines in the loop regions of HTR sequences supports the formation of G-quadruplex structures. Any substitution of A-for-G increases the melting temperature, while the introduction of several substitutions was found to facilitate the coexistence of several conformers in the presence of potassium. The systematic introduction of these substitutions finally leads to the formation of sequences which occur in the Tetrahymena telomere. In addition, similar sequences were also found in the human genome. These findings raise an interesting point. Why does the Tetrahymena telomere require sequences which can adopt such highly stable G-quadruplex structures? In general, very stable G-quadruplexes are usually a source of problems in cells during the life cycle of an organism. The THR sequence is more polymorphic than HTR; it forms two different monomeric and one dimeric conformers as has been shown here and in our previous studies [14]. Our analysis focused on sequences consisting of four G-runs without any overhanging nucleotides at both termini; this type of arrangement is not an ideal model for extrapolation to natural telomeric repeats which typically consist of tens to thousands repeats.

Our results demonstrate that all HTR derivatives including THR can be converted from antiparallel to parallel folds in the presence of potassium and PEG-200. ICD spectra indicate that the binding mode of TO with THR in the presence of KCl might be different from those observed for HTR derivatives, and this is a finding which could also be important for other molecules recognizing the THR structure in nature. It suggests that the structure of THR shows some structural features which are different from those of HTR and HTR derivatives in the presence of potassium. Confirmation of the biological significance of this fact remains an open topic.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Slovak Research and Development Agency under Contracts nos. APVV-0280-11 and APVV-0029-16, European Cooperation in Science and Technology (COST CM1406), Slovak Grant Agency (1/0131/16 and 002UPŠ-4/2015), and internal university grants (VVGS-PF-2017-251 and VVGS-2016-259). The authors thank G. Cowper for critical reading and correction of the manuscript.

References

[1] K. W. Lim, V. C. M. Ng, N. Martín-Pintado, B. Heddi, and A. T. Phan, "Structure of the human telomere in Na⁺ solution: An antiparallel (2+2) G-quadruplex scaffold reveals additional


On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids

Review Article

Ewan K. S. McRae,1 Evan P. Booy,1 Gay Pauline Padilla-Meier,1 and Sean A. McKenna1,2,3

1Department of Chemistry, University of Manitoba, Winnipeg, MB, Canada
2Department of Biochemistry & Medical Genetics, University of Manitoba, Winnipeg, MB, Canada
3Manitoba Institute for Materials, University of Manitoba, Winnipeg, MB, Canada

Correspondence should be addressed to Sean A. McKenna; sean.mckenna@umanitoba.ca

Received 28 July 2017; Accepted 8 October 2017; Published 9 November 2017

Guanine quadruplexes (G4s) are four-stranded secondary structures of nucleic acids which are stabilized by noncanonical hydrogen bonding systems between the nitrogenous bases as well as extensive base stacking, or pi-pi, interactions. Formation of these structures in either genomic DNA or cellular RNA has the potential to affect cell biology in many facets including telomere maintenance, transcription, alternate splicing, and translation. Consequently, G4s have become therapeutic targets and several small molecule compounds have been developed which can bind such structures, yet little is known about how G4s interact with their native protein binding partners. This review focuses on the recognition of G4s by proteins and small peptides, comparing the modes of recognition that have thus far been observed. Emphasis will be placed on the information that has been gained through high-resolution crystallographic and NMR structures of G4/peptide complexes as well as biochemical investigations of binding specificity. By understanding the molecular features that lead to specificity of G4 binding by native proteins, we will be better equipped to target protein/G4 interactions for therapeutic purposes.

1. Introduction

Nucleic acid polymers that are rich in guanines have the potential to form a four-stranded structure called a guanine quadruplex (G4). When four guanines are arranged in a plane they are stabilized by hydrogen bonds between the Watson-Crick and Hoogsteen faces of adjacent guanines. These guanine tetrads have extensive surface area for pi-pi base stacking with other tetrads which contributes greatly to the thermodynamic stability of the G4 structure. All guanine G4s share this core feature of stacked tetrads of guanines, which is likely a common feature contributing to the recognition of G4s by proteins. Formation of G4 is dependent upon the presence of monovalent cations that relieve negative electrostatic charge repulsion from the O6 of guanines concentrated in the center of the tetrad. Positioning of monovalent cations (typically sodium or potassium) in the center of the G4, between two stacked tetrads, relieves the repulsive effect of the negative electrostatic potential [1, 2].

G4s can have greatly different topology based on the relative orientation of the phosphodiestere backbone connecting the runs of guanines. If all the strands are parallel with respect to the 5’ and 3’ orientation of the ribose sugars, the loop sections between the runs of guanine must connect from the top to the bottom of the G4 in what is known as a propeller orientation (Figure 1(a)). The glycosidic bond angles in parallel G4s are typically all in the anti conformation, making this the preferred strand orientation for RNA G4s as it better accommodates the 2’ hydroxyl group. Similarly, locked nucleic acids (LNA) that use 2’-O,4’-C methylene linkages to force a C3’-endo sugar pucker will also strongly favor the all anti glycosidic bond conformation afforded by a parallel G4s [3]. For this reason, LNA have been utilized when forcing a parallel G4 topology is desired. For antiparallel strands the loops are referred to as lateral if connecting adjacent strands and diagonal if connecting strands opposite them; these create a mixture of syn and anti glycosidic bond
angles and are the structures most often observed in DNA G4s [4] (Figure 1(b)).

Much like other nucleic acid secondary structures, distinct grooves created by the sugar-phosphate backbone can be observed in the space filling models of G4 structures (Figure 1(c)). These grooves as well as the large surface of the tetrads on either end of the G4 and the loops themselves represent the three feasible interaction sites for protein binding partners. Lateral and diagonal loops in antiparallel G4 leave the grooves accessible but would restrict access to the tetrad face. Propeller-like loops as seen in the parallel G4 pass directly over the grooves, allowing for intramolecular interactions of loop bases with the grooves, thereby obscuring the grooves but leaving the tetrad face more exposed for protein or ligand binding (Figure 1). Amongst the myriad of small molecule G4-binding ligands that have been synthesized one can find examples of ligands binding grooves, loop sequences, and tetrad faces of G4s. In some cases, binding of these ligands has been shown to stabilize, destabilize, and even alter the conformation of the G4 [5–7]. It would not be unreasonable to hypothesize that as our knowledge of protein-G4 interactions increases we will have examples of proteins that do the same.

Bioinformatic analyses of the human genome identified some ∼350,000 unique, putative, G4 forming sites (PQS) [8] and high-throughput sequencing methods have identified a further 450,000 [9]. Interestingly, there appears to be an evolutionary pressure against PQS in protein coding regions and an enrichment of PQS in other noncoding regions when compared to pseudo-randomly generated DNA sequences. This enrichment of PQS in noncoding regions supports the idea that G4s may be key regulatory elements. Moreover, there is a statistical enrichment of PQS in oncogene promoter regions, some of which have been confirmed to exist in vivo by G4 specific cross-linking studies [10] and other biochemical investigations [11–14]. It has also become clear over the past decade that G4s play an important role in maintenance of telomere length [15–17] and ribosome biogenesis [18], which are disregulated in ∼80% of cancers [19]. Furthermore, the expansion repeats of guanine rich regions in mRNA have been linked to two common forms of neurological disorders, Fragile X mental retardation [20], and familial Amyotrophic Lateral Sclerosis (ALS) [21].

There exists some controversy as to whether G4 can form in vivo and while significant effort has gone towards testing this question experimentally, observing the structure of nucleic acids in vivo is a monumental task that is prone to biasing dependent on the method used to determine structure. One method that is used to visualize unique motifs in cells is immunofluorescence. The Balasubramanian group has developed antibodies to selectively target both DNA [22, 23] and RNA G4 [24] and successfully used them to visualize these structures in vivo. Furthermore, one could reason that the physiological effects of quadruplex-specific small molecules on cells are evidence of quadruplex formation in vivo. In these cases, the arguments can be made that the antibody or small molecule is merely inducing the formation of G4 and that the structures do not exist in the absence of these probes or that the specificity of the antibodies/drugs has changed in a cellular context and that they are no longer binding to G4 but exerting an effect through interactions with other structures. A recent study using high-throughput sequencing and in vivo chemical modifications has indicated that RNA G4s are globally unfolded in some mammalian cell lines [25]. A similar critique can be made that the chemicals added for modification of nucleic acids could be altering

Figure 1: Side view of the human telomere guanine quadruplex in parallel (a) and antiparallel (b) conformations. Space filling models (c) for parallel (left) and antiparallel (right) quadruplexes. PDBs 1KF1 (a) and 143D (b). Guanines shown in green, thymines shown in blue, adenines shown in red, and the phosphoribose backbone shown in grey.
their structure in vivo or that the concentration of chemical reagent added results in nonspecific effects. Regardless of being able to prove the existence of G4s in vivo, many proteins have highly conserved G4 specific binding domains that would likely not have persisted in eukaryotes if they had no biological function. One way to probe this purpose is by disruption of the protein-G4 interaction.

Targeting G4–protein interactions with small molecule drugs has potential to treat a multitude of diseases [17, 26, 27]. Currently there is one such compound in clinical trials for cancer treatment, that is, CX-3543 or Quarfloxin. While originally selected for its interaction with the MYC G4, CX-3543 has been shown to localize to cell nuclei where it inhibits Pol I transcription by blocking the interaction between nucleolin and G4s in rDNA [18]. As well as being targets of small molecule therapeutics, G4s themselves have been employed as therapeutics to target specific proteins. One such example is thrombin binding DNA aptamers that inhibit thrombin-induced platelet aggregation and clot-bound thrombin; these have potential for use as anticoagulants during various surgeries and are currently in clinical trials [28]. Although thrombin is not by nature a G4-binding protein, G4 aptamers display remarkable affinity and specificity for regions of thrombin that normally interact with other proteins.

To better understand the mechanism of G4-mediated cellular processes and how small molecules interfere with them, it is imperative to understand how proteins recognize their endogenous G4 binding partners. While many G4-binding proteins have been identified and validated [29], there is relatively little information on how they specifically recognize G4s. Typical approaches to study this include altering some of the features of either the G4 substrate or protein, subsequently assessing how these mutations affect the interaction, and inferring from this their role in G4 recognition. This method is relatively simple and inexpensive to perform and can yield useful information; however it is not always trivial to show that such mutations are directly mediating the interaction and not indirectly affecting the interaction by altering the conformation of the mutated species. Small-angle X-ray scattering is a solution-based technique that can be used to determine low-resolution models of protein and nucleic acid. Though it does not provide enough resolution to distinguish individual atoms, it can provide insight into the shape and orientation of species in a protein nucleic acid complex that can be complementary to other data obtained from mutation analysis and/or other higher-resolution structural methods such as NMR and X-ray crystallography. While there are currently over 250 high-resolution NMR and X-ray structures of G4s deposited on the Research Collaboratory for Structural Bioinformatics (RCSB) website and nucleic acid database (NDB), comparatively few high-resolution structures of protein interacting with G4 have been solved. The most extensively studied interaction is between G4 containing DNA aptamers and thrombin, on which a number of structures have been deposited (PDB: 1HAP, 1HAO, 3QLP, 4DII, 4DH1, 5CMX, 4I7Y, 4LZ1, 4LZ4, 5LUW, 5LUY, 5EW1, and 5EW2). More recently the structure of a peptide from the prion protein bound by an RNA G4 aptamer has been solved (2R8K, 2RU7). Currently, to our knowledge, only 3 high-resolution structures have been solved showing G4-peptide interactions involving proteins that interact with G4 as part of their normal cellular function (2LA5, 5DE5, and 2N21). Herein we will discuss the insights gained from these structures as well as some key mutational analysis studies and look at the benefits, as well as drawbacks, of these approaches.

## 2. Thrombin Binding by G4 DNA Aptamers

The development of thrombin binding DNA aptamers for their anticoagulant properties has been ongoing since the early 1990s; although thrombin is not naturally a G4-binding protein, the 15-base DNA G4 (TBA), developed through SELEX, can tightly bind thrombin and inhibit fibrin clot formation [31, 32]. This nucleic acid sequence forms an antiparallel G4 with two guanine tetrads connected by three lateral loops: one face of the G4 is covered by two TT loops and the other by a TTA loop. Initial high-resolution structures of TBA (1HIP, 1HAO) showed that the fibrinogen binding site of thrombin (exosite I) was likely interacting with the loops of TBA; however, due to lack of electron density in the loop regions it was unclear whether this interaction was with the TT loops or TTA loop [33]. 15 years later 3 structures (3QLP, 4DII, and 4DH1) were produced from the lab of Filomena Sica that definitively show the TT loops acting as a pincer that bind either side of the protruding region of exosite I utilizing multiple polar and hydrophobic contacts [34, 35].

Many variants of the aptamer have been made that enhance the stability of, affinity for, and inhibition of thrombin. One such potent second-generation thrombin aptamer is the HD22-27mer [36] for which a high-resolution X-ray structure (4IY7) of the aptamer-thrombin complex has been solved [37] (Figure 2(a)). HD22-27mer forms a mixed duplex-G4 structure where the duplex is directly enchainced to the G4. The HD22-27mer forms an unusual pseudo-G4 topology with four loops, where the first loop connects two guanines in the same tetrad and the remaining loops form a more typical lateral antiparallel loop pattern. The 2nd and 4th of these loops interact with each other through Watson-Crick A-T hydrogen bonds and cap one side of the G4, while the intratetrad loop and the 3rd loop also form A-T hydrogen bonds and cap the opposite side of the G4. In contrast to the structure of thrombin and TBA, protein recognition of HD22-27mer involves the extended double stranded region as well as a bulged-out thymine and the pseudo-G4 core and appears to involve exosite II and not exosite I. 2G0 is the last guanine before the duplex region and forms three polar contacts with Arginine 93, two through ribose hydroxyl groups and one through its phosphate group. Loops 2 and 4 of the G4 contribute to the stability of the complex by forming multiple hydrophobic interactions with thrombin but, unlike with TBA, only thymine 9 has multiple polar contacts with exosite II.

Recently a 31-nucleotide third-generation thrombin aptamer, RE31, was crystalized in complex with thrombin [38] (Figure 2(b)). Like the HD22-27mer, RE31 has an extended duplex region as well as a two-tetrad G4. Unlike...
HD22-27mer the duplex region of RE31 exhibits continuous base stacking with the G4 rather and the duplex region does not appear to be important for binding to thrombin. Despite the presence of a G4-duplex junction, the binding of RE31 to thrombin is much more similar to TBA than HD22-27mer; both molecules use the TT loops to interact with exosite I through hydrogen bonds and hydrophobic interactions.

Despite the persistence of a two-tetrad G4 in these thrombin binding aptamers, interactions with the G4 core are not observed in any of the structures. Instead, the main site of interaction, which is common between all the structures mentioned above, is with the second and fourth loop of the G4. While TBA and RE31 use mainly polar contacts and some ancillary hydrophobic interactions to interact with exosite I of thrombin, the G4 loops of HD22-27mer use mostly hydrophobic interactions to bind exosite II. This common mechanism of binding to different sites of thrombin suggests that the G4 acts as a scaffold that presents the single stranded loops to the binding pocket on the protein.

3. Interaction between the RGG Motif and G4 Structures

Motifs rich in arginine and glycine have been known to play functional roles in numerous physiological processes, usually involving nucleic acid interactions, for many decades [39]. Glycine is the most flexible of the amino acid side chains in terms of its accessible dihedral angles; thus sequences rich in glycine can sample a large variety of conformations allowing them to be very promiscuous binding partners. Arginine is a positively charged amino acid with delocalized pi electrons. This allows for favorable electrostatic interactions with the phosphate backbone of nucleic acids, as well as potential base stacking with the nitrogenous bases and hydrogen bonding in as many as three different directions via its guanidinium moiety. Many of the currently known G4-binding proteins have RGG/RG motifs [29, 39], some of which have been shown to be critical to G4 binding and recognition, including the translated in sarcoma/fused in sarcoma (TLS/FUS) [40], nucleolin (NCL) [41], Ewing's sarcoma protein (EWS) [42], Epstein Barr virus nuclear antigen I EBNA1 [43], DDX21 [44], and fragile X mental retardation protein (FMRP) [45].

Since the most obvious structurally unique feature of a G4 is the tetrad face, one might expect the arginine of the RGG motifs to contribute to base stacking interactions with the terminal guanine tetrads. However, there is only indirect evidence that the tetrad face plays a role in RGG mediated interactions. Many small molecule G4-binding ligands have been shown to interact tightly with the tetrad face via pi-pi stacking interactions and two of such molecules, BRACO-19 and CX-3543, have been shown to disrupt interactions with G4-binding proteins (EBNA1) [43] and NCL [18], which contain the RGG motif. While this indirectly implicates competition with the protein for the tetrad face of the G4, these compounds also bind electrostatically to the grooves and loops of the G4 [46, 47] possibly preventing protein-groove interactions or reorienting the loop structure to disrupt protein binding in such a manner.

3.1. Recognition of Phosphoribose Backbone of G4 Loops by RGG/RG Domains. Biochemical investigations in the lab of Takanori Oyoshi have revealed that the RGG domains of the EWS and FUS/TLS proteins are critical for G4 binding and do so through interactions with the phosphoribose backbone of the loop regions [48–50]. The EWS protein's RGG3 domain has high affinity for specific RNA and DNA G4 that is dependent on the arginines in this region [42]. Interestingly this binding seems to be dependent on the presence of loops.
of at least 2 nucleotides and higher affinity is observed with longer loops [50]. Replacing the loop sequences with abasic deoxy-ribose backbones did not affect this affinity, indicating that the interaction was with the backbone itself and not the nucleotide bases. In contrast, both the backbone and bases of the TT loops in TBAs are involved in thrombin binding and thus replacing the TT loop residues of TBAs with abasic backbone reduces the affinity for thrombin as well as the orientation of the aptamer on the surface of thrombin [51–53]. Furthermore, an intermolecular G4 with single stranded overhangs did not demonstrate the same affinity, indicating the structure of the backbone in the G4 loops was important for recognition by the RGG3 domain. While the initial strand orientation of the G4 did not affect binding, RGG3 was shown by circular dichroism to convert the topology of DNA G4 from antiparallel to parallel. Thus, it seems that the phosphoribose backbone structure in the propeller type loop conformation is being specifically recognized by the RGG3 domain of EWS [50].

Mutation of the three phenylalanine residues in the RGG domain of FUS/TLS to tyrosine eliminates the binding to the DNA G4 while retaining affinity for the RNA G4 [48]. Interestingly, when the DNA G4 loops were replaced with an abasic phosphoribose backbone, the binding capability was restored. This indicates that the tyrosine residue specifically recognizes the $2'$ hydroxyl group of ribose sugars in the loop region and not the guanine tetrads. RNA G4 with no loops or locked nucleic acid loops (which obscure the $2'$ hydroxyl) abrogate the binding, consistent with the notion that the phosphoribose backbone structure in the propeller type loop conformation is being specifically recognized by the RGG3 domain of EWS [49].

These mutational studies with EWS and FUS/TLS demonstrate that RGG domains, like in the case of thrombin binding aptamers, can recognize the loops of G4 structures. Both RGG domains preferentially interact with G4-containing longer loops as opposed to those lacking loops. Furthermore, the RGG domains of EWS and FUS/TLS need only the phosphoribose backbone to be present in the loops for full affinity, whereas thrombin interacts with both the backbone and bases of TBA loops. This indicates that the binding of G4 by EWS and FUS/TLS is not sequence specific and that the position of the backbone of the loops is likely key to specific recognition by EWS and FUS/TLS.

3.2. G4 Binding by the RGGGGR Peptide of the FMRP Protein.

Currently the only high-resolution structures of an RGG domain of FUS/TLS to tyrosine has been previously shown to retain specificity and affinity for G4 structures and not other RNA that the full-length FMRP protein interacts with, indicating it is responsible for G4 recognition by FMRP [56]. Discrepancies between the structures are attributed to flexible regions within the peptide and loops of the RNA (Figure 3(a)). The structure shows a G4 of parallel strand orientation with 3 guanine tetrads and an unusual mixed base tetrad (GUAU) before the duplex junction. While there are no direct interactions with the guanine tetrads, Arg15, which was shown to be indispensable for binding, is seen to exhibit a cation-pi stacking interaction with A17, which is part of the mixed base tetrad. In addition to this stacking interaction, Arg15 is observed in a hydrogen bond with the Hoogsteen face of G7, the first base in the double stranded region (Figure 3(b)).

Arg10, another indispensable residue, rests in between the first and second bases of the duplex region (C30 and G31) and exhibits a cation-pi interaction with C30 and forms multiple hydrogen bonds with G31 and the phosphoribose backbone. Mutation of these first two base pairs has been shown to reduce the affinity of the peptide for SCI tenfold [56]. The two arginines that make contact with these bases act as an anchor at the base of the G4 stabilizing the junction from four- to two-stranded RNA, a feature that was exemplified by RNase digestion experiments.

The four glycines between Arg10 and Arg15 form a tight beta-turn motif and they make many important contacts with the RNA. Gly11 appears to be indispensable (based on mutations made to the peptide), as it coordinates hydrogen bonds from its backbone NH and CO to the carbonyl of G4 and the NH$_2$ group of C5 in the duplex region. Gly14 makes the second contact with the G4 region via a hydrogen bond from its backbone amide to the ribose sugar of the uridine in the mixed tetrad (Figure 3(b)).

Unlike the thrombin binding aptamers and RGG domains of EWS and FUS/TLS, the SCI RNA-FMRP interactions do not involve the loop bases at all; the high-resolution structures and mutational analysis show that the important interactions are with the duplex region adjacent to the G4 as well as the mixed base tetrad. From the space filling model (Figure 3(c)) it appears that the RGGGGR peptide is occupying a groove created by the tetraplex to duplex junction; thus its mode of recognition appears to be a mixture of van der Waals space filling and sequence specific interactions mediated by the placement of arginines in this cavity.

4. G4 Recognition of the Tetrad Face by the DHX36 Specific Motif

The protein DHX36 is a helicase enzyme that has the capability to bind and unwind RNA G4s with great specificity [57,58]. A region of 13 amino acids that is unique to DHX36 is evolutionarily conserved amongst higher eukaryotes and is both necessary and sufficient for binding of G4s [59].
Figure 3: RGGGGR peptide bound to SC1 RNA quadruplex-duplex junction. Comparison of peptide conformations between X-ray (green) and NMR (yellow) structures shown in (a). Hydrogen bonding pattern between peptide and nucleic acids shown in (b). Space filling model shown in (c), with peptide shown as a magenta ribbon. PDBs 5DEA (X-ray) and 2LA5 (NMR).

Low-resolution small-angle X-ray scattering has shown that the G4 recognition motif is oriented towards the tetrad face (Figure 4(c)) \cite{30, 60} and mutations to the loop sequence of an endogenous G4 binding partner resulted in a reduced affinity, indicating that loop conformation may also be important for recognition of G4 by this motif \cite{61}. A high-resolution NMR structure from the lab of Anh Tuấn Phan showing a small 18-amino-acid peptide bound to a synthetic DNA G4 TT(GGGT)x4 (2N21) supports this mode of binding \cite{62}. It is worth noting that a similar mode of binding has been observed between an RNA aptamer and a 16-amino-acid peptide from the N-terminus of the prion protein PrP \cite{63, 64}.

Since DHX36 has significant preference for the parallel versus antiparallel loop orientation, the loop sequences were limited to a single thymine residue which promotes formation of the parallel species. The structure shows the propeller type loop arrangement typical of parallel G4s with the peptide forming a short alpha-helical region followed by a turn that covers the entire tetrad face. The two aromatic residues (Trp13 and Tyr14) do not make any contacts with the DNA but are purported to stabilize the L shape of the peptide. Four residues, namely, Gly9, Gly13, Ile12, and Ala17, are found at the tetrad-peptide interface close enough to permit CH-pi interactions (Figure 4(a)).

Positively charged residues Lys8, Arg10, and Lys19 are each in close enough proximity to the grooves of the G4 to form electrostatic interactions which could help stabilize the observed interaction (Figures 4(b)/4(d)); however, previous mutation studies have shown Lys8 and Arg10 to be dispensable in the context of the full-length protein \cite{59}. The observed mode of binding explains DHX36’s preference for RNA G4s; since RNA G4s form almost exclusively the parallel strand orientation with propeller type loops, they are expected to have exposed tetrad faces. Binding to a large exposed hydrophobic surface like the tetrad face provides a significant increase in entropy through desolvation and is thus a highly favorable interaction. DNA G4s which have antiparallel or hybrid type loop structures can avoid this entropy penalty through interactions between the loop bases and tetrad face which bring the polar sugar-phosphate backbone over the top of the G4 and shield the hydrophobic tetrad face.

Limitations of NMR and X-Ray Crystallography. NMR and X-ray crystallography are two techniques that have many
strengths and limitations which complement each other and make the use of both techniques the ideal method to investigate biological structures of interest. Where X-ray crystallography will provide a single image of the exact structure in the crystal lattice, NMR will show many models of possible structures which all fit the data equally well. NMR is a solution-based technique which can view the dynamics of molecules on a physiological time scale, whereas X-ray crystallography views, usually, the most thermodynamically stable form of a system. When analyzing X-ray and NMR structures, it is important to remember that they provide only a snapshot of a dynamic process, usually under nonphysiological conditions. Insights gleaned from these snapshots should be tested with mutational studies that use the information from the structure to test its validity.

The main pitfall of NMR for structural biology is the size restriction on the molecule being investigated. Especially for proteins and nucleic acids, past a certain length, there are many overlapping signals from different regions of the biomolecule resulting in broad, indiscrète peaks, which makes separation and identification of which signal came from which atom an immense challenge. Therefore, often, only small peptides are examined by this method. The disadvantage of this approach is that you cannot ensure that the small peptide is not folded as it would be in the context of the full-length protein. For example, the 18-amino-acid peptide from DHX36 is only 2% of the size of the full 1008-amino-acid protein. While the peptide does demonstrate significant affinity for quadruplex, the full-length protein binds much more tightly [59]. Thus any structural insights gained from examining the peptide are, at best, only a piece of the full picture. Furthermore, it has been observed from the structures of the FMRP/Sc1 complex that recognition of G4s is not just from the G4 itself but the junction regions...
and adjacent double or single stranded regions; thus it is also preferable to use a larger oligonucleotide which better mimics biological RNA/DNA species. The problem with doing so is that such complexes can exceed the size limit for structural determinations by NMR and that such large complexes are usually difficult to crystallize, often requiring nonnative salt conditions or stabilizing antibodies. The disadvantage to such crystallization approaches is the difficulty to prove reliability of any single structure obtained in terms of physiological relevance.

One of the advantages of NMR approaches is the ability to identify disordered regions. Even without relaxation dispersion experiments, the variability in models obtained by NMR can hint at which regions are disordered. X-ray crystallography, on the other hand, traps disordered regions into a single stable conformation which can confuse data interpretation and cause misleading results. An example of this trapping of flexible regions is in the X-ray structure of the FMRP RGG domain and the Sc1 RNA. The NMR data on the complex, which was obtained first, had poorly defined loop regions, hinting at conformational flexibility. Fortuitously two different loop conformations were captured in a single asymmetric unit in the crystal conformation supporting the notion that the loop region samples multiple conformations. Had there been only one complex in the unit cell, there would be less certainty in the dynamics of the loop nucleotides.

Another consideration for studying G4s by either NMR or crystallography techniques is that they require high biomolecule concentrations that can lead to artifacts being observed in both the crystal lattice and the NMR data. For example, the first structure of a natural G4 and protein in the same crystal cell unit was observed for the Oxytrichia nova telomere end binding protein (OnTEBP). OnTEBP has been shown to bind to the single stranded repeats at the end of telomeres (TTTTGGGG in *Oxytrichia nova*) and protect them from DNA damage repair mechanisms which recognize breaks in dsDNA [65]. Horvath and Schultz [66] were studying the binding of the single stranded repeat (GGGGTTTTGGGG) to the OnTEBP region and upon analyzing their X-ray crystallography data observed diffraction from a DNA quadruplex which appeared to be bound to OnTEBP (1JB7). Three symmetry related OnTEBP proteins appeared in the unit cell each bound to the ssDNA and each contributing to the binding of a single DNA quadruplex formed from a dimer of their single stranded substrate. Interactions with one OnTEBP protein were with a major groove in the G4 backbone and featured both electrostatic and hydrogen bonds to the deoxyribose sugar groups. The other two OnTEBP proteins bound via the TTTT loop sequences via electrostatic interactions, hydrogen bonds, and pi-pi as well as CH-pi stacking interactions. The authors recognized in their paper that the interactions were likely due to crystal packing interactions and may not be physiologically relevant, which serves as an example of why one must take care when analyzing data which was acquired under nonphysiological conditions.

Conditions such as low pH, high salt, and high sample concentration are often used to obtain either better diffracting crystals or more discrete NMR bands. Figure I exemplifies the importance of careful consideration of the solution in which structural information is obtained as well as the method of obtaining such information. Two completely different topologies are observed for the same DNA sequence dependent upon which cation is stabilizing the quadruplex. The presence of Na+ stabilizes the antiparallel [67, 68] form whereas K+ enables crystallization of the parallel [69] form.

5. Conclusions

Mutational analysis and high-resolution NMR and X-ray studies have provided invaluable insights into the structures of biological molecules; this in turn has expounded the molecular mechanism of many important biological processes. With these structure function relationships, we can rationally approach how to alter these processes and modify them to fix pathologies. Although the recognition of G4s by proteins is far from being understood, several trends have emerged amongst the data collected to date. It is apparent that the loop structure of G4s and the junction regions adjacent to them present a unique molecular landscape for recognition by proteins. So far this appears to be the most common mode of recognition. Thrombin binding aptamers appear to canonically bind thrombin using the thymine bases of two lateral loops like a pincer to latch onto thrombin. The RGG domains of EWS and FUS/TLS also appear to interact mainly with the loops of G4s; the mode of recognition in this case is likely from the placement of the phosphoribose backbone that connects the strands of the G4 since the bases themselves have been shown to be dispensable. In the case of FMRP, the RGGGGR peptide stabilizes the transition from G4 to duplex by filling the junction between them with base stacking and Hoogsteen type hydrogen bonds with the double stranded region.

Mutations made to the FUS/TLS protein and its G4 substrate have shown that preferential recognition of RNA over DNA G4 can be mediated by 2'OH interactions. It will be interesting to see how other G4-binding proteins are affected by such mutations as those performed with FUS/TLS protein. So far, DHX36 is the only G4-binding protein that has been shown to directly interact with the open guanine-tetrad face of the G4 and this satisfactorily explains its preference for RNA G4s, since the faces of antiparallel G4s are typically obscured by the loops. However, since much of the affinity for G4 is lost when only examining the DHX36 specific motif peptide, there are likely other binding interfaces on the G4 which contribute to the full-length protein's binding affinity.

Currently most G4-binding small molecules interact directly with the guanine-tetrad face. This would, obviously, directly compete with DHX36 binding but has also been shown to prevent binding of other RGG domain containing G4-binding proteins, making specific targeting and mechanistic elucidation of their effects challenging [18, 43, 47]. An alternate approach to designing compounds to bind G4s would be to design compounds that mimic their unique loop structure. Since loop recognition seems to be key to the specificity of many G4-binding proteins, creating small molecules to compete with this interaction may result in more
specific compounds that could be used in a more targeted manner to affect cellular pathways.

Conflicts of Interest
The authors declare that there are no conflicts of interest regarding the publication of this paper.

References


Research Article

On the Helical Structure of Guanosine 5'-Monophosphate Formed at pH 5: Is It Left- or Right-Handed?

Gang Wu,1 Irene C. M. Kwan,1 Zhimin Yan,2 Yining Huang,2 and Eric Ye3

1Department of Chemistry, Queen's University, 90 Bader Lane, Kingston, ON, Canada K7L 3N6
2Department of Chemistry, Western University, London, ON, Canada N6A 5B7
3Department of Chemistry, University of Ottawa, Ottawa, ON, Canada K1N 6N5

Correspondence should be addressed to Gang Wu; wugang@queensu.ca

Received 24 July 2017; Accepted 11 October 2017; Published 2 November 2017

Academic Editor: Gary Parkinson

Copyright © 2017 Gang Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Gel formation of guanosine 5'-monophosphate (5'-GMP) under slightly acidic conditions (e.g., pH 5) was first discovered by Bang in 1910 [1]. However, it was not until 50 years later that the structural basis of such 5'-GMP gel was examined. In 1962, Gellert et al. [2] used X-ray fiber diffraction data to show that different GMP isomers form different helical structures. For 3'-GMP gel, the helical structure is formed by successive stacking of planar hydrogen-bonded guanine tetramers now known as the G-quartets on top of each other. For 5'-GMP gel formed at pH 5, in contrast, the planar (disc-like) G-quartet is broken at one side forming a lock-washer-like structure which is further hydrogen bonded into a continuous helix; see Figure 1. In 1975, Sasisekharan et al. [3] further investigated the helical structure formed by 5'-GMP at pH 5 (i.e., 5'-GMP gel) and reported atomic coordinates for a left-handed 15/4 helix model. However, these authors also noted in the paper that "[b]ecause the helix is not constrained by a continuous covalently bonded backbone, both right- and left-handed helices of the 15/4 model can be constructed. Although stereochemically quite different, they are both acceptable. ...For arbitrary reasons, we have selected for detailed examination a left-handed helix...." Therefore, it appears that, on the basis of the original fiber X-ray diffraction data alone, there is no particular reason to favor a left-handed helix over a right-handed one. However, this arbitrary choice of the helical structure has been overlooked in the literature so that, even in classic treatises of nucleic acid structures such as that by Saenger [4], this helix is described as left-handed. It is also clear from the study of Sasisekharan et al. [3] that whether the acidic 5'-GMP helix is left- or right-handed depends critically on the sugar pucker conformation. That is, a C2'-endo sugar pucker would favor a left-handed helix but a C3'-endo conformation would result in a right-handed helix. However, because 5'-GMP gels are difficult to study with common spectroscopic techniques, the question regarding its exact helical structure has never been fully addressed.

In 2009, we used solution NMR techniques to obtain structural details of the helix formed by 5'-GMP at pH 8 [5]. As shown in Figure 1, the physical appearance of the 5'-GMP solution depends critically on the pH. At pH 8, the 5'-GMP solution appears as a normal liquid, whereas,
at pH 5, it becomes a gel. The key findings of our earlier study of the 5'-GMP helix formed at pH 8 are as follows. First, the central structural motif of the helix is the disc-like G₄. Second, the 5'-GMP molecules take alternating C2'-endo and C3'-endo sugar pucker conformation along the helical strand. Third, the helix is right-handed. Fourth, the central channel of the helix is filled with Na⁺ ions each being sandwiched between two disc-like G₄ s. In contrast, as Sasisekharan et al. [3] proposed, the central structural motif of the helix formed by 5'-GMP at pH 5 is a lock-washer-like G₄ structure, as illustrated in Figure 1. However, other details about this helix are not known. Since 5'-GMP forms gel at pH 5, conventional solution NMR techniques are not applicable. In this work, we applied solid-state NMR and IR methods to obtain structural details about the helical structure formed by 5'-GMP at pH 5 (5'-GMP gels). In particular, we set out to address key questions concerning sugar pucker conformation, phosphate-phosphate interaction, phosphate-base interaction, and metal ion binding environment around the helical structure.

2. Experimental Sections

Hydrated disodium salt of guanosine 5'-monophosphate (purity > 99%) was obtained from Sigma-Aldrich (Ontario, Canada). The 5'-GMP gel sample was prepared by acidifying 1.0 M Na₂(5'-GMP) aqueous solution to pH 5 with acetic acid. Before the solid-state NMR experiments, the gel was gently dried with a stream of N₂. The 1D ¹H MAS and 2D ¹H double quantum (DQ) NMR experiments were performed at 21.1 T with a Bruker 1.3 mm HX probe with a sample spinning frequency of 62.5 kHz. The back-to-back (BABA) recoupling sequence [7] was used for the ¹H DQ experiments with the excitation time being set to one rotor period. The recycle time employed was 8 s. The 2D ¹H → ³¹P HETCOR experiments were performed at 21.1 T with a Bruker 2.4-mm MAS probe. The sample spinning frequency was 33 kHz. Contact times from 0.5 to 2.0 ms were employed. Solid-state ¹³C CP/MAS NMR experiments were performed at 14.1 and 21.1 T. All ¹³C chemical shifts were referenced to that of TMS by setting the ¹³C signal of a solid sample of tetrakis (trimethylsilyl) silane (TKS) to 3.50 ppm. Solid-state ³¹P NMR experiments were performed on a Bruker Avance-600 spectrometer operating at 242.96 MHz for ³¹P. All ³¹P chemical shifts were referenced to 85% H₃PO₄ (aq). Solid-state ²³Na NMR experiments were performed on a Bruker Avance-500 spectrometer operating at 132.72 MHz for ²³Na nuclei with the following parameters: sample spinning, 10 kHz; ¹H decoupling, 65 kHz; recycle time, 10 s; 64 transients. All ²³Na chemical shifts were referenced to NaCl (aq) at δ = 0.0 ppm by setting the ²³Na signal of NaCl(s) to 7.21 ppm. ²³Na[³¹P] REDOR experiments using the original version of the pulse sequence [8] were performed on a Varian/Chemagnetics Infinity-Plus 400 WB
spectrometer operating at a magnetic field strength of 9.4 T. The $^{31}$P and $^{23}$Na resonance frequencies at this field strength are 161.72 and 105.67 MHz, respectively. All MAS spectra were acquired using a Varian/Chemagnetics T3 4-mm triple-tuned MAS probe. Typical RF power levels corresponded to 180° pulse lengths of 70 and 78.4 μs for $^{23}$Na and $^{31}$P nuclei, respectively. A total of 512 transients were accumulated for each REDOR measurement. The sample spinning rate was kept constant at 10000 ± 2 Hz. The recycle delay was 0.2 s.

3. Results and Discussion

To assess the basic self-assembled structure of 5'-GMP gel formed at pH 5, we first obtained its $^1$H solid-state NMR spectra under very fast MAS conditions at an ultrahigh magnetic field, 21.1 T (900 MHz for $^1$H). For comparison, we also reported the corresponding $^1$H NMR spectra for crystalline Na$_5$(5'-GMP)-7H$_2$O (orthorhombic). As seen in Figure 2, for the acidic 5'-GMP gel sample, the N$^1$H and N$^1$H$^4$ signals appear at about 10.6 ppm, suggesting that both protons are involved in strong hydrogen bonding. The DQ signals connecting N$^1$H$^4$ and H$^8$ provide the most direct evidence for G$_4$ formation, although this feature alone cannot reliably distinguish between the planar disk-G$_4$ and lock-washer-G$_4$ motifs (vide infra). Interestingly, two N$^2$H signals are seen for Na$_5$(5'-GMP)-7H$_2$O (orthorhombic). This observation is consistent with the crystal structure of the compound where there are two distinct 5'-GMP molecules in the asymmetric unit [9]. This doubling of the signals is more evident in the $^{13}$C CP/MAS spectrum of Na$_5$(5'-GMP)-7H$_2$O (orthorhombic) (see Figure S1 in the Supporting Information, available online at https://doi.org/10.1155/2017/6798759). Furthermore, for Na$_5$(5'-GMP)-7H$_2$O (orthorhombic), the N$^2$H signals appear at about 13.5 ppm, whereas the corresponding N$^1$H signals are between 4 and 6 ppm. These observed $^1$H chemical shifts are in agreement with the crystal structure of Na$_5$(5'-GMP)-7H$_2$O (orthorhombic) which shows that N$^1$H forms a strong hydrogen bond with "O-P (the two N...O distances are 2.76 and 2.79 Å) and the N$^2$H$_4$ groups are only weakly hydrogen bonded to water molecules (two N...O$_W$ distances are 2.91 and 2.95 Å) [9]. It is interesting to note that both acidic 5'-GMP gel and Na$_5$(5'-GMP)-7H$_2$O exhibit DQ signals between H$^8$ and H$^5$, consistent with the guanine base being in the anti-conformation. Now, while the $^1$H solid-state NMR data confirm G4 formation for the acidic 5'-GMP gel, they provide no information about the sense of the helix.

As mentioned earlier, on the basis of modeling performed by Sasisekharan et al. [3], whether the acidic 5'-GMP helix is left- or right-handed depends critically on the sugar pucker conformation. To answer this question, we utilized a well-established approach in using $^{13}$C chemical shifts of the sugar carbons as a means of determining the sugar pucker conformation. In particular, Harbison and coworkers [10, 11] showed that, for RNA nucleosides and nucleotides, one can combine the $^{13}$C chemical shifts observed for the ribose moiety into the following two canonical coordinates:

$$
can1 = 0.179\delta (C1') - 0.225\delta (C4') - 0.0585\delta (C5'),
$$

$$
can2 = -0.0605 [\delta (C2') + \delta (C3')] - 0.0556\delta (C4') - 0.0524\delta (C5').
$$

(1)

Then any data point appearing in the can1-can2 plot can be used to determine the sugar pucker conformation (can1 > -6.77 for C3'-endo and can1 < -6.77 for C2'-endo) as well as the exocyclic γ-torsion angle (can2 < -16.82 for gt and can2 > -16.82 for gg). Later, Ohlenschläger et al. [12] applied this approach to analyze a total of 429 known RNA structures and showed that the reliability of this approach for purine nucleotides is 93-94% (see Figure S2 in the Supporting Information).

Figure 3(a) shows the solid-state $^{13}$C CP/MAS NMR spectrum of acidic 5'-GMP gel where the observed $^{13}$C chemical shifts for sugar carbons C1', C2', C3', C4', and C5' are 879.76, 69.3, 82.2, and 62.8 ppm, respectively. This assignment was further confirmed by DFT calculations on the $^{13}$C chemical shifts for a model 5'-GMP molecule. These values yield can1 = -6.43 and can2 = -16.67 for the acidic 5'-GMP gel. Now the fact that can1 > -6.77 and can2 > -16.82 for the acidic 5'-GMP gel strongly suggests that the sugar pucker conformation is exclusively C3'-endo with the exocyclic γ-torsion angle being in the gg conformation [10–12]; see Figure S2 in the Supporting Information. These canonical coordinates are quite different from those for Na$_5$(5'-GMP)-7H$_2$O (orthorhombic) and 5'-GMP self-assembly formed at pH 8; also see Figure S2 in the Supporting Information. This new information about the C3'-endo sugar pucker conformation means that the continuous helix of the acidic 5'-GMP gel is right-handed. To further confirm the C3'-endo sugar pucker conformation determined above, we recorded FTIR spectra for three 5'-GMP samples. Some time ago, Tajmir-Riahi [13] showed that the P-O-5'-ribose stretch frequency can be used as the signature for the sugar pucker conformation for guanylic acid and its salts: 800 cm$^{-1}$ for C3'-endo and 820 cm$^{-1}$ for C2'-endo. As seen from Figure 3(b), the acidic 5'-GMP gel sample indeed displays a peak at 800 cm$^{-1}$, confirming the aforementioned C3'-endo sugar pucker conformation. In comparison, the FTIR spectrum of the 5'-GMP self-assembly formed at pH 8 exhibits both 800 and 820 cm$^{-1}$ peaks of equal intensity. This is in agreement with the earlier observation that the helical structure of 5'-GMP formed at pH 8 consists of alternating C3'-endo and C2'-endo sugar pucker conformation [5]. For Na$_5$(5'-GMP)-7H$_2$O (orthorhombic), the observation of a peak at 820 cm$^{-1}$ is in agreement with its crystal structure where the ribose is in the C2'-endo conformation [9]. Therefore, the FTIR data shown in Figure 3(b) are fully consistent with the results on sugar pucker conformation obtained from the $^{13}$C chemical shift analysis. Now, combining the C3'-endo sugar pucker conformation with the helical parameters reported by Sasisekharan et al. [3], we can readily build a right-handed 15/4 helix model; see Figure S3 and Table S1 in the Supporting Information for atomic coordinates.

Since metal ion binding is an integral part of G-quadruplex formation [14–18], we further investigated how
Figure 2: 2D $^1$H DQ NMR spectra of (a) dried 5'-GMP gel formed at pH 5 and (b) Na$_2$(5'-GMP)-7H$_2$O (orthorhombic). The corresponding 1D $^1$H NMR spectra are shown at the top. All $^1$H NMR spectra were obtained under the MAS condition with a sample spinning frequency of 62.5 kHz at 21.1 T.

Figure 3: (a) $^{13}$C CP/MAS NMR spectrum of dried 5'-GMP gel formed at pH 5. (b) The signature region of IR spectra revealing sugar pucker conformation for 5'-GMP gel (pH 5), 5'-GMP (pH 8), and Na$_2$(5'-GMP)-7H$_2$O (orthorhombic).

Na$^+$ ions are bound to the acidic 5'-GMP helical structure. Figure 4 shows the solid-state $^{23}$Na NMR spectra obtained for the acidic 5'-GMP gel as well as for a neutral 5'-GMP self-assembly sample for comparison. The two $^{23}$Na NMR signals observed for the acidic 5'-GMP gel can be readily assigned: the sharp signal at 7.2 ppm is due to fully hydrated Na$^+$ ions and the signal centered at −5.0 ppm is from phosphate-bound Na$^+$ ions. To further confirm the phosphate-bound nature of the signal at −5.0 ppm, we performed $^{23}$Na($^{31}$P) rotational-echo double resonance (REDOR) [8] experiments. As shown in Figure 5, the $^{23}$Na($^{31}$P) REDOR results obtained for the acidic 5'-GMP gel are quite similar to those for neutral 5'-GMP and for double-stranded calf thymus DNA in the dry state (A-form) [19]. Thus the $^{23}$Na($^{31}$P) REDOR results confirmed unambiguously that the $^{23}$Na NMR signal at −5 ppm arises from Na$^+$ ions bound to the phosphate group. The most striking feature in the $^{23}$Na NMR spectrum of acidic 5'-GMP gel is the absence of any signal at ca. −18 ppm, which is the established spectral signature for Na$^+$ ions residing inside a G-quadruplex channel [20–22]. This observation immediately suggests that there is no Na$^+$ ion inside the central channel of the continuous helix formed by 5'-GMP at pH 5! This aspect of the helix, though totally unexpected, can be readily understood on the basis of our structural model. As seen from Figure 6, when a disc-like G$_4$ is twisted into a lock-washer-like G$_4$, the size of the central
cavity surrounded by carbonyl oxygen atoms is significantly reduced. As a result, Na\(^+\) ions can no longer fit into this cavity. The diameter of the central channel is reduced by nearly 50% for the acidic 5'-GMP helix compared with that of the neutral 5'-GMP helix, as clearly seen from the top view of the channel shown in Figure 6. This observation is consistent with the fact that 5'-GMP gel formation at pH 5 is not sensitive to the nature of monovalent cations (Na\(^+\), K\(^+\), or NH\(_4\)\(^+\)) present in solution. It is also worth noting that the helical structure of acidic 5'-GMP gel is remarkably similar to that found for 8-oxoguanosine reported recently by Giorgi et al. [23], despite the very different hydrogen bonding schemes in these two systems. Here we further comment on the role that the central cations play in G-quadruplex systems consisting planar disc-like G-quartets. While it is commonly accepted that the central cation is to reduce the repulsion between carbonyl oxygen atoms from G-quartets, it is important to point out that it is primarily the repulsions between carbonyl oxygen atoms from adjacent planar G-quartets, not from within the same G-quartet, which requires further stabilization from a cation. The main evidence for this view is the fact that whereas a cation-free or "empty" G-quartet was observed [24], an "empty" G-octamer has never been reported. Now when the helix is made of lock-washer-like G\(_4\), there is no longer any repulsion between carbonyl oxygen atoms along the helical axis, thus making it unnecessary to have a cation inside the central channel.

Now let us turn attention to the phosphate group in the acidic 5'-GMP helix. Since the phosphate group of 5'-GMP has a \(pK_{a_2}\) of 7.5, it is doubly charged at pH 8 but only singly charged at pH 5. We discovered in an earlier study [5] that two types of phosphate groups are present in the 5'-GMP helix formed at pH 8 and they are possibly bridged by a Na\(^+\) ion. For the 5'-GMP helix formed at pH 5, our model suggests that singly charged phosphate groups form a continuous hydrogen-bonded chain along the helical "strand" (i.e., \(\cdots\text{HO-P}^-\cdots\text{O}^-\cdots\text{HO-P}\cdots\cdots\)). This type of hydrogen bond chains are commonly observed in the crystal structures of ammonium hydrogen alkylphosphates [25]. Because of this strong hydrogen bonding interaction, the P\(\cdots\)P distance is significantly shorter in the acidic 5'-GMP helix, 5.2 \(\text{Å}\), than in the neutral 5'-GMP helix (6.7 and 7.2 \(\text{Å}\)) [5]. The solid-state
$^{31}$P NMR spectrum of acidic 5'-GMP gel exhibits a sharp peak at 1.3 ppm (vide infra), suggesting that all phosphate groups are equivalent. This is in contrast to the situation seen in the neutral 5'-GMP helix where two different phosphate groups are present with the $^{31}$P chemical shifts being 3.7 and 5.2 ppm [5]. Another important structural feature in the acidic 5'-GMP helix is the possible formation of a phosphate-base hydrogen bond, as first noted by Sasisekharan et al. [3]. In particular, the $i$th phosphate group can be hydrogen-bonded to the exocyclic amino group of the $(i+3)$th guanine base (i.e., N$_2$-H$^B$-O=P) along the helical strand. In our model, the N$_2$⋯O(P) distance is ca. 2.82 Å.

To search for further spectroscopic evidence for the aforementioned two types of hydrogen bonding interactions involving the phosphate group (i.e., ⋅⋅⋅HO-P$_i$-O⋅⋅⋅HO-P$_{i+1}$-O⋅⋅⋅ and N$_2$-H$^B$⋯O=P), we performed 2D $^1$H → $^{31}$P HETCOR experiments. As seen in Figure 7(a), at a short contact time of 0.5 ms, two cross peaks were observed. The weaker cross peak with $\delta(^1H)$ of 4.1 ppm clearly arises from the short contacts between the phosphorus atom and H$^{5'}$,$^5''$ (2.65 and 3.04 Å), as illustrated in Figure 7(b). The stronger $^1$H-$^{31}$P cross peak with $\delta(^1H)$ of 10.5 ppm is an interesting discovery, because we have already attributed, in the earlier discussion, N$_1$H and N$_3$H$^A$ to this overlapping signal. Now we see that a third signal, which displays the shortest contact with the P atom, also appears in this $^1$H chemical shift region. This new signal must be due to the P-OH group (the H-P distance is ca. 2.24 Å in our model); see Figure 7(b). As seen in Figure 7(a), a new cross peak corresponding to the N$_1$H$^B$ group emerges at a longer contact time (2 ms). This is consistent with our model where the P atom is predicted to be 3.14 Å away from N$_1$H$^B$, due to the formation of a N$_2$-H$^B$⋯O=P hydrogen bond. This hydrogen bond further explains why the $^1$H chemical shift of N$_2$-H$^B$, ca. 8 ppm, is considerably higher than those seen in the neutral 5'-GMP helix, 5.12 and 4.29 ppm [5].

4. Conclusion

In this work, we have obtained new structural details about the helical structure formed by 5'-GMP at pH 5. Contrary to the common assumption, we showed that this helix is composed of 5'-GMP molecules exclusively in C$^3'$-endo sugar pucker conformation and consequently is right-handed. In addition, we found that the central channel of the helix is free of Na$^+$ ions. In many aspects, this helix is drastically different from the one formed by 5'-GMP at pH 8. Remarkably, two different helices can form by the same molecule at just slightly different pH values. Of course, at pH 5 and 8, the charge state of the phosphate group would be different. The present study has provided another example where mononucleotides can self-associate into a helix in the absence of phosphodiester bonds. The solid-state NMR strategies demonstrated in this study can be applied to similar gels formed by other nucleosides and nucleotides.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The
authors thank Andy Kalevar for his contributions at the early stage of this research.

References

Review Article

In What Ways Do Synthetic Nucleotides and Natural Base Lesions Alter the Structural Stability of G-Quadruplex Nucleic Acids?

Janos Sagi

Rimstone Laboratory, RLI, Carlsbad, CA 92010, USA

Correspondence should be addressed to Janos Sagi; jans@rimstonelab.com

Received 22 June 2017; Accepted 15 August 2017; Published 18 October 2017

Academic Editor: Tom Marsh

Copyright © 2017 Janos Sagi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Synthetic analogs of natural nucleotides have long been utilized for structural studies of canonical and noncanonical nucleic acids, including the extensively investigated polymorphic G-quadruplexes (GQs). Dependence on the sequence and nucleotide modifications of the folding landscape of GQs has been reviewed by several recent studies. Here, an overview is compiled on the thermodynamic stability of the modified GQ folds and on how the stereochemical preferences of more than 70 synthetic and natural derivatives of nucleotides substituting for natural ones determine the stability as well as the conformation. Groups of nucleotide analogs only stabilize or only destabilize the GQ, while the majority of analogs alter the GQ stability in both ways. This depends on the preferred syn or anti N-glycosidic linkage of the modified building blocks, the position of substitution, and the folding architecture of the native GQ. Natural base lesions and epigenetic modifications of GQs explored so far also stabilize or destabilize the GQ assemblies. Learning the effect of synthetic nucleotide analogs on the stability of GQs can assist in engineering a required stable GQ topology, and exploring the in vitro action of the single and clustered natural base damage on GQ architectures may provide indications for the cellular events.

1. Introduction

In the course of the last two decades, substantial worldwide research activity has been devoted to the elucidation of the structural and biochemical properties of the non-B, polymorphic GQ structures, both in vitro and in vivo. Many laboratories used synthetic nucleotide building blocks, which were reviewed by several publications [1–5]. Studies have also focused on such site-specific incorporations of nucleotide analogs that could prompt the modified oligonucleotides to fold into specific GQ topologies, as summarized in [6–8]. In addition to folding topology, the fundamental property of GQ structures is their thermodynamic stability. Multiple aspects of the effect of mutation by natural nucleotides on GQ stability are well known [2–5, 9–11]. Properties of GQs built from peptide nucleic acid (PNA) units [12, 13] and of a few other special modifications, such as the (R)-1-O-(pyren-1-ylmethyl)glycerol (intercalating nucleic acid, INA) or (R)-1-O-[4-(1-pyrenylethynyl)-phenylmethyl]glycerol (twisted intercalating nucleic acid, TINA) have also been analyzed [4, 14]; therefore, these modifications are not discussed here. Much less has been summed up on the effect of synthetic constituents on the thermal and thermodynamic stability of the modified GQs, which is thus the subject of the present review. Other important current topics are the effect of single and multiple natural base lesions on GQ stability as well as the epigenetic modifications occurring in GQs, the present status of which is also reviewed here up to midsummer 2017.

Single or double-helical guanine-rich (G-rich) natural or synthetic DNA and RNA sequences can, under appropriate physical (cations, cosolvents) or cellular (e.g., negative superhelical stress) conditions, convert into hairpin, triple-helical, and four-stranded GQ structures. A GQ comprises two or more G-tetrads, which are connected by loops of one or more nucleotides. The planar G-tetrads are formed through Hoogsteen-type circular double hydrogen bonds (Figure 1), and the cations are located either in the central cavity of the G-tetrad or in the spaces between the tetrads. The GQ fold...
The effect of the synthetic nucleotide analogs is discussed here according to their stabilizing and/or destabilizing action. Those, around a dozen analogs that, under the specific conditions studied, were found to only stabilize or not affect the stability beyond the experimental error of determination of a specific GQ fold, make up one group of the analogs. Majority of the synthetic nucleotides studied so far that either stabilize or destabilize the GQ constitute another, the largest group, and those analogs that only destabilized a GQ belong to the third group of synthetic nucleotide analogs. Similarly, the natural base lesions were also grouped according to their stabilizing or destabilizing effects.

Stability of a GQ structure can be most properly characterized by its thermodynamic parameters communicated by the free energy change value, Δ\(G^\circ\), at various temperatures, such as Δ\(G^\circ\)\(_{37}\), in kcal/mol. Thermodynamic parameters can be determined by the model-dependent method, such as the “two-state” method using temperature-dependent melting profiles by UV or CD absorption spectroscopy [22, 23], or, more exactly, by the model-independent differential scanning calorimetry (DSC) method [24, 25]. DSC is certainly the more accurate method as it explores the details of the thermal transition contrary to the integrating absorption methods (although not all folding-unfolding transitions fit the two-state requirement, several laboratories neglect it and use the two-state absorption method). Both methods provide the half-way thermal transition temperature, \(T_m\) or \(T_{1/2}\), \(T_{1/2}\) has been used when the annealing (refolding) is slower than the unfolding, which results in hysteresis of the thermal profile. Since most studies do not publish free energy change values but all publish \(T_m\) or \(T_{1/2}\) values, the best way to compare the stabilities of various GQ structures is to use the thermal stability values. Δ\(G^\circ\) values will be mentioned herein in selected cases. Figure 3 shows typical thermal melting profiles as determined by UV absorption spectroscopy at 296 nm [26].

2. Synthetic Nucleotide Analogs

That Stabilized or Barely Impacted the Stability of a GQ

Few synthetic analogs of natural nucleotides are currently known, which only stabilize a GQ independently of their syn or anti isomeric conformation and/or the position of the nucleotide they replaced in the potential GQ-forming oligonucleotide. Similarly, a limited number of analogs have so far been described that do not affect the thermodynamic stability of GQs beyond the experimental error of determination. It is important to note that these effects were observed in a given GQ system; the nucleotide analogs may induce the opposite effect in other, so far not-studied GQs.

N2-Guanine Derivatives. The C2-aminogroup of dG (Figure 4, 1) is one of two positions of dG whose substitutions do not interfere with the circular double H-bonding of a G-tetrad [27, 28]. Substitution of N2 does not change the glycosidic torsion angle of the anti dG. The 2-amino group of guanine substituted with hydrophobic groups can enhance the thermodynamic stability of a GQ via disrupting the water sphere around the core. Water depletion is known to stabilize a GQ (and destabilize duplex DNAs) [29, 30]. The TGGGAG oligonucleotide DNA modified at the 3‘- and 5‘-ends with 3,4-dibenzyloxybenzyl groups (R-95288, Hotoda oligonucleotide), which forms a parallel tetramolecular GQ (Figure 2, 6), has anti-HIV-1 activity through binding to...
The intramolecular GQ architectures adopted by the 21-mer htel-21 \( G_3(TTAG)_3 \) oligodeoxynucleotide under various solution conditions, topologies 1–5. The basket-type antiparallel with 5'–to-3' edgewise-diagonal-edgewise loops, forming in Na\(^+\) (topology 1); the chair-type antiparallel with three edgewise loops (topology 2); the (3 + 1 strands) hybrid-1 with propeller-like, edgewise, and edgewise loops (topology 3); the hybrid-2 with two edgewise loops followed by a propeller-type loop (topology 4), both hybrids forming in K\(^+\) solution; the parallel topology with all three loops in propeller-like configuration, as found in crystals containing K\(^+\) ions (topology 5); tetramolecular GQ assembling from four TGGGT DNA chains, [TG\(\_\)T]\(_4\) (topology 6). Bottom panel: the two-tetrad, chair-type intramolecular antiparallel topology, folding from the 15-mer TBA oligodeoxynucleotide sequence shown below the topology. The \( \text{syn} \) dG nucleotides are indicated with underlined letters in italics. For references, see, for example, [8]. Shaded squares represent the G-tetrads displayed in Figure 1.

gp120 [31]. dGs of this oligodeoxynucleotide had been substituted with N2-methyl-2'-deoxyguanosine (m\(^2\)dG). The m\(^2\)dG modifications enhanced the stability of the parent GQ and the modified oligonucleotides showed a 2-fold higher anti-HIV activity than the R-95288. The GQs formed from T[m\(^2\)G]GGAG and TGGm\(^2\)GAG had higher \( T_m \) values, 70°C, than that assembled from TGGGAG, \( T_m \) of 60°C. On the other hand, the 3'-terminal substitution with m\(^2\)dG had no effect on the GQ stability due to the poor stacking of the terminal base. The increased stabilities were explained by improved base stacking originating from the methylation of NH\(_2\) of G2. The GQ of T[m\(^2\)G]\(_3\)AG with three contiguous m\(^2\)dG residues had the highest \( T_m \), which was over 80°C. The alkylation of N2G seems to be a promising way to improve the stability of parallel GQs, as long as the modification does not sterically hinder the formation of the GQs [32]. Substitution of N2 of G was also mentioned in connection with the TBA GQ (Figure 2, 7), however, only in connection with the thrombin-binding activity: attaching of a benzyl group to N2 of the \( \text{anti} \) N-glycosidic dGs (Figure 4, 1) of positions G6 and G11 and a 1-naphthylmethyl group into the N2 of G6 of the TBA increased the thrombin inhibitory activity. The 1-naphthylmethyl group substitution at the N2 of G6 showed about a 60% increase in activity both \textit{in vitro} and \textit{in vivo}.  

**Figure 2:** Folding arrangements of the most often cited GQ structures.
Figure 3: CD spectra of different GQ topologies and a series of thermal unfolding profiles. (a) shows the CD spectra of the antiparallel basket architecture of htel-21 GQ recorded in 100 mM Na\(^+\) solution and characterized by positive Cotton effects near 295 nm and 240 nm and a negative one around 260 nm (solid line), as well as the polymorphic form adopted by the same oligodeoxynucleotide in 100 mM K\(^+\), which is a mixture of K\(^+\)-stabilized antiparallel and the hybrid forms, with the positive ellipticity near 290 nm, a strong shoulder around 270 nm, and a negative one close to 240 nm (broken line). (b) displays, for comparison, the CD spectra of the same K\(^+\)-stabilized mixture of the htel-21 GQ (dotted line) and the mixture of just the hybrid forms with two positive peaks of close to equal heights at 290 and 260 nm (dashed line), all determined at low (50–100 \(\mu\)M) DNA strand concentrations, as well as the majority parallel forms adopted at/above 2 mM of the strands (solid line) (see references in [8]). (c) illustrates normalized thermal unfolding profiles of 8-oxoguanine containing htel-21 GQs [26].

Substitutions on N2 of other G residues of the sequence had a little effect or decreased the activity [33]. The latest example for stabilization of GQ structures by N2-dG substitutions comes from using htel GQs. Lech and Phan [34] reported on the incorporation of N2-methyl, N2-hexylamino, and the N2-benzyl derivatives of dG into an htel-24 DNA GQ, the TT(htel-21)A. The substitutions stabilized the G-quadruplex by 2 to 13 \(^\circ\)C per modification.

Isoguanine (iG). 2-Oxo-6-aminopurine, isoguanine (Figure 4, 2), has been thoroughly studied in GQ models, first by Seela and coworkers [35] with \(T_4iG_T_4\) that formed a tetramolecular parallel GQ, which contained iG-tetrads held together by circular double hydrogen bonds. The \(T_4iG_T_4\) also proved to be a more stable structure than the parent \(T_4G_T_4\) in buffered 1 M NaCl, 10 mM MgCl\(_2\), pH 7 [36]. The \(T_4iG_T_4\) also formed parallel GQ with iG-tetrads [37]. The iG formed not only tetrads but also pentads, giving rise to pentaplexes, such as \(T_4iG_T_5\), and the pentad planes were held together, as the tetrads, by circular double H-bonds [38–40]. The iG-pentaplexes were found to bind and activate the Fe(III)-heme complex towards peroxidase activity, comparably to GQs and contrary to iGQs [41]. In the intramolecular GQ of TBA, a single iG residue was found to enhance the binding of GQ to the human R-thrombin, as compared to the parent GQ aptamer. The most active aptamer was built...
Figure 4: Structures of the *anti* N2-substituted-dG (1); isoguanine (2); 3-halo-3-deazaguanine (3); syn 8-bromo-dA (4); 8-propynyladenine (5); *anti* 5-fluoro-dU (6); *anti* 5-bromo-dU (7); 5-aminouracil (8); and oligo(ethylene glycol) substituted *anti* 5-propynyl-dU (9).

from GGTTGGT iG TGGTTGG, and its approximate binding enhancement was 2-fold over the unmodified aptamer. The authors supposed that iG-GQs tended to form and denature more easily than the corresponding structures bearing G. Whether this would mean increased or decreased thermodynamic stability is not known [21].

3-Deazaguanine (c3G). Quantum chemical methods-based models showed that 3-halo-3-deazaguanines (Figure 4, 3) replacing guanines in two- or three-tetrad GQs result in substantial energy gain leading to elevated stability for the GQs, which mainly originates from increased π-π stacking interactions [42].

8-Bromo-, 8-Methyl-, and 8-Propynyladenines. In the tetramolecular GQ-forming TAGGGT and AGGGT oligodeoxynucleotides, the adenines have been replaced by 8-bromadenine (br8dA), 8-propynyladenine (py8dA) (Figure 4, 4, 5) [43–45], and 8-methyladenine (m8A) [46, 47]. All these A-modified oligonucleotides formed the same parallel-type tetramolecular GQs as the two unmodified oligonucleotides did. While py8A decreased the stability of duplex DNAs, it markedly elevated the thermal stability of the tetramolecular structures, assumingly, due to a prevalent glycosidic syn conformation [43, 45]. br8A also increased the stability of the parent structures [44]. The GQ analogs of TAGGGT did not contain T, A, or 8-substituted-A tetrads, whereas the GQ of AGGGT formed an A-tetrad at the 5’-end. An m8dA tetrad was also formed; it was an all-syn m8dA tetrad that had only a minor effect on \( T_m \) of the parent structure.

5-Fluorouracil (f5U). Thymines of the TBA oligodeoxynucleotide were replaced, one by one, with 5-fluoro-2'-deoxyuridines (Figure 4, 6) by Virgilio et al. [48]. The 5-fluoro substitution for the 5-methyl group elevated the thermal stability of the GQ in each of the six cases studied, mostly by 1–3°C; however, in positions 4 and 13, f5U enhanced the stability by 11 and 10°C, respectively. These two GQs also showed improved anticoagulant activity. Thymines bases at the T4 and T13 loop positions stack to the tetrad residues [18–20]. The high \( T_m \) values may indicate that the fluoro atom enhanced the strength of stacking. Thymines in other loop positions do
not interact with the GQ core. The modified GQs adopted the same type of antiparallel topology as the wild-type did, based on the CD spectra.

5-Bromo-, 5-Amino-, and 5-Hydroxyuracils. Tetrad thymines had been replaced by the synthetic 5-bromouracil (br\(^5\)U) and 5-aminouracil (n\(^5\)U) (Figure 4, 7, 8) that substituted for the central T of TGG\(_T\)GGT. The oligonucleotides modified by these dU-analogs formed tetramolecular parallel GQs, as the unmodified did, with all residues having anti glycosidic conformation. Order of thermal stability was n\(^5\)U > br\(^5\)U > T-. containing GQs. The [TGG\(_n\)\(^5\)UGGT]\(_4\) GQ showed a biphasic melting profile, which was explained by noncooperative melting. The n\(^5\)dU tetrad was observed having circular double HBonds and this provided higher stability than T or br\(^5\)U [49]. Using Density Functional Theory calculations and ESI-MS measurements, Paragi et al. [50] hypothesized that 5-amino- and especially 5-hydroxyuracils can form four pyrimidine and also mixed tetrads with guanine or xanthine, where tetrads have similar energy and cover almost exactly the same area as the full guanine or xanthine tetrads.

5-Propynyluracil and Its OEG and PEG Derivatives. Based on their previous findings [51] Tateishi-Karimata and coworkers [52] described in a recent paper the stabilizing effect of 5-propynyl-2'-deoxyuridine on the TBA GQ when it was substituted for T4 and T7 of the loops. T base at position 4 stacks to the G-tetrad, while the T7 does not [18–20]. The propynyl-uracil substitution of T4 enhanced the stability of the native TBA GQ (\(T_m\): 50.7 C, \(\Delta G_{-25}^{\circ}\): −3.5 kcal/mol) up to 61.7 C and −5.4 kcal/mol, whereas the nonstacking did not change it: 50.9 C and −3.6 kcal/mol, in buffered 100 mM KCl. The propynyluracil was further derivatized with ethylene glycols (Figure 4, 9). Oligo- and polyethylene glycols (OEG and PEG) are used as crowding agents \textit{in vitro} to mimic the intracellular environments. Both destabilize the dsDNA but stabilize the GQs [30]. Various lengths and numbers of ethylene glycols were attached to the 5-(1-propynyl)-tethered dUs and substituted for the T4, T7, and T13 of the TBA GQ to learn whether these glycols interact with the GQ. Based on thermodynamic and NMR spectral analyses, they have found that the glycols do interact with the bases of the GQ, and the octaethylene (OEG8) and longer-chain glycols elevated the thermodynamic stability of the parent GQ beyond the stability increase induced by the propynyl tether alone. For instance, the \(T_m\) and \(\Delta G_{-25}^{\circ}\) values of 50.7 C and −3.5 kcal/mol for the TBA GQ were increased to 64.3 C and −7.1 kcal/mol and 55.2 C and −4.4 kcal/mol for the propynyl-OEG(8)-dU at positions 4 and 7, respectively. The results suggested that PEG molecules interact with tetrad and loop bases via CH-π and lone pair-π interactions [52].

5-(Benzo[h][1,3]dioxole-2-yl)uracil. Tanpure and Srivatsan [53] reported that the 5-(benzo[h][1,3]dioxole-2-yl)uracil (Figure 5, I) can be used to monitor the formation of GQs from htel DNA and

*Figure 5: Structures of 5-(benzofuran-2-yl)uracil (1); threose dinucleotide (2); R- (3) and S-isomer (4) of acyclic-uracil derivatives and the n-propyl spacer (5).*
RNA sequences. The uracil probe exhibited 4- to 9-fold fluorescence intensity enhancements in the GQ compared with its duplex structure. The effect depended on the GQ topology and on the position of the modification. The probe minimally perturbed the GQ topology and stability and could distinguish the antiparallel, hybrid, and parallel topologies of DNA and RNA GQs from the corresponding duplexes. UV thermal melting profiles showed that the unmodified and the modified GQs had similar $T_m$ values, within ±2°C, even in different ionic conditions.

**Threose Nucleotide (TNA).** The sugar analog α-L-threofuranosyl nucleotide (TNA for threose nucleic acid) (Figure 5, 2) has been examined if the oligonucleotides containing this shorter internucleotide building unit can also form GQs [54]. Despite the modified backbone repeat unit, which is one atom shorter than that found in DNA and RNA, TNA could self-assemble into stable tetramolecular GQ structures that are similar in thermal stability to the 2'-deoxyribose-containing $[\text{TG}_4\text{T}]_4$. Unlike DNA, the TNA GQs formed equally well in either Na⁺ or K⁺ ions [54].

**Acyclic-Uracil Derivatives.** The R-stereoisomer of an acyclic uracil containing a cyclopentane ring fused at positions 5 and 6 of the uracil ring (R-c, Figure 5, 3) had been introduced by Borbone et al. [55] into position T7 or T12 (TG'T and the TT loops) of the TBA oligodeoxynucleotide. The resulting two aptamers folded into the typical antiparallel chair-like GQ structures formed also by the native TBA in K⁺ buffer (Figure 2). The acyclic residue increased the thermal stability of the resulting GQs with respect to the TBA GQ, by 1°C and 4°C, respectively. Anticoagulant activity of the activity the TBA-T7 was more potent than the TBAs in prolonging clotting time [55]. Later, the research group incorporated both the S and the R stereoisomers (S-c and R-c; Figure 5, 3, 4) into the same TBA sequence into positions 3, 7, and 12. Positions 3 and 12 are located in the two TT loops, and position 7 is located in the TG'T loop. At these locations, both isomers of the acyclic-T derivative enhanced the thermal stability of the resulting GQs and all adopted the same antiparallel topology as the parent TBA GQ did. $T_m$ of the unmodified GQ, 50°C, was raised to 54–55°C by the R-derivatives and to 51–56°C by the S-derivatives [56] (it is not known how these derivatives would influence the thermal stability in the other three T positions of the loops).

**n-Propyl Spacer.** A three-carbon spacer (Figure 5, 5) replaced the base-sugar residues of T3, T7, and T12 loop positions of the TBA GQ. The flexible abasic sites increased $T_m$ of the unmodified TBA GQ (49°C) by 6°C at each of the three positions and did not alter the antiparallel chair folding topology of the unmodified GQ. The modifications improved the thrombin clotting time [57].

**Other Stabilizing Analogs.** A novel fluorescent TBA analog has been created by conjugation with a dansyl probe at the 3’-end and a β-cyclodextrin residue at the 5’-end. The bis-conjugated TBA analog could fold into the chair-like, antiparallel GQ structure in K⁺, typical of the native TBA GQ, and retained its thrombin-binding properties as well. The terminal appendages only marginally affected the conformational features of the TBA GQ. Folding into GQ was associated with a net fluorescence enhancement due to encapsulation of the 3’-dansyl group into the apolar cavity of the β-cyclodextrin at the 5’-end. This novel analog of the TBA GQ demonstrated its potential as a diagnostic tool for thrombin recognition and also provided a useful basis for the design of suitable aptamer-based devices for therapeutic applications, that is, allowing simultaneous detection and inhibition or modulation of the thrombin activity. The terminal double conjugation increased $T_m$ of the native TBA by about 10°C [58].

The short TG₄T oligodeoxynucleotide forms a stable tetramolecular parallel-stranded GQ in K⁺ solution. The 3’- and/or the 5’-ends of four TG₄T molecules have been linked together by a non-nucleotide-type, tetra-end-linker (TEL) by Oliviero et al. [59]. The TEL-[TG₄T]₄ architectures were able to form parallel GQs regardless of the TEL size and the structural orientation of the oligonucleotide branches. The TEL-[TG₄T]₄ structures that had longer TEL were more thermostable than those with the shorter TEL.

A perylene group, a five-membered fused aromatic structure, has been linked to the nonbonded oxygen of the phosphodiester linkage between dG-dG, dG-T, and T-T dinucleotides, and a single perylene-modified dimer was introduced in the TBA and the 20-mer (TGGGT)₄ oligonucleotides to develop fluorescence anisotropy GQ probes. Single modifications showed little effect on GQ conformation and the thermostability increased in both K⁺ and Pb⁺⁺ if the perylene moiety was not located between the dG stacks. With some modified sequences, $T_m$ of the resulting GQ increased by 32°C. In most cases, the GQs were more stable in Pb⁺⁺ than in K⁺ with $T_m$ increasing up to 26°C. The large stabilization was probably due to the interaction of perylene with the G-quartet plane and a more compact structure was induced by Pb⁺⁺. CD spectra of the (TGGGT)₄ GQ and the modified analogs showed parallel topology in the presence of K⁺ with a positive peak near 265 nm and a negative one near 240 nm. The TBA and its analogs displayed positive peaks near 245 and 295 nm and a negative one near 265 nm, typical of the antiparallel GQ structures. In the presence of Pb⁺⁺, the positive peaks of 265 nm and 295 nm of the structures were red-shifted, indicating the Pb⁺⁺-stabilized antiparallel GQs in both scaffolds [60].

Non-nucleoside-type linkers, such as propanediol, octanediol, or hexaethylene-glycol, that replaced the whole TTA loops connecting the four GGG tracts of an htel GQ induced the formation of parallel GQs, like the single-TEL was more thermostable than the wild-type TTA loop-containing 5’-TGGGT-TAG₃-TAG₃-TAG₃-Q GQ did ($T_m$ 69°C), while the wild-type htel GQ. In 40 mM K⁺, all loop-replaced htel analogs melted at higher temperatures than the wild-type TTA loop-containing 5’-F-G₃-TAG₃-TAG₃-TAG₃-Q GQ did ($T_m$ 69°C), while the wild-type htel GQ. The 5’-end and a β-cyclodextrin residue at the 5’-end. The bis-conjugated TBA analog could fold into the chair-like, antiparallel GQ structure in K⁺, typical of the native TBA GQ, and

The short TG₄T oligodeoxynucleotide forms a stable tetramolecular parallel-stranded GQ in K⁺ solution. The 3’- and/or the 5’-ends of four TG₄T molecules have been linked together by a non-nucleotide-type, tetra-end-linker (TEL) by Oliviero et al. [59]. The TEL-[TG₄T]₄ architectures were able to form parallel GQs regardless of the TEL size and the structural orientation of the oligonucleotide branches. The TEL-[TG₄T]₄ structures that had longer TEL were more thermostable than those with the shorter TEL.

A perylene group, a five-membered fused aromatic structure, has been linked to the nonbonded oxygen of the phosphodiester linkage between dG-dG, dG-T, and T-T dinucleotides, and a single perylene-modified dimer was introduced in the TBA and the 20-mer (TGGGT)₄ oligonucleotides to develop fluorescence anisotropy GQ probes. Single modifications showed little effect on GQ conformation and the thermostability increased in both K⁺ and Pb⁺⁺ if the perylene moiety was not located between the dG stacks. With some modified sequences, $T_m$ of the resulting GQ increased by 32°C. In most cases, the GQs were more stable in Pb⁺⁺ than in K⁺ with $T_m$ increasing up to 26°C. The large stabilization was probably due to the interaction of perylene with the G-quartet plane and a more compact structure was induced by Pb⁺⁺. CD spectra of the (TGGGT)₄ GQ and the modified analogs showed parallel topology in the presence of K⁺ with a positive peak near 265 nm and a negative one near 240 nm. The TBA and its analogs displayed positive peaks near 245 and 295 nm and a negative one near 265 nm, typical of the antiparallel GQ structures. In the presence of Pb⁺⁺, the positive peaks of 265 nm and 295 nm of the structures were red-shifted, indicating the Pb⁺⁺-stabilized antiparallel GQs in both scaffolds [60].

Non-nucleoside-type linkers, such as propanediol, octanediol, or hexaethylene-glycol, that replaced the whole TTA loops connecting the four GGG tracts of an htel GQ induced the formation of parallel GQs, like the single-TEL was more thermostable than the wild-type TTA loop-containing 5’-TGGGT-TAG₃-TAG₃-TAG₃-Q GQ did ($T_m$ 69°C), while the wild-type htel GQ. In 40 mM K⁺, all loop-replaced htel analogs melted at higher temperatures than the wild-type TTA loop-containing 5’-F-G₃-TAG₃-TAG₃-TAG₃-Q GQ did ($T_m$ 69°C), while the wild-type htel GQ.
3. Stabilization or Destabilization of a GQ by the Same Synthetic Nucleotide Analog

Depending on the folding topology, GQs contain nucleosides of anti and syn N-glycosidic linkages, which connect the base and sugar moieties. In most cases, the parallel GQs of any molecularity contain all anti nucleosides, and the antiparallel and hybrid architectures (Figure 2) comprise both types of natural nucleosides. Although the thermodynamic variance between the two conformation states is small, in the range of 1-2 kcal/mol, forcing a nucleoside from syn to anti or vice versa does affect the thermodynamic stability of a GQ and generally leads to destabilization. In some cases, it also induces topological rearrangement of the GQ. In this way, all 8-substituted 2'-deoxyguanosines whose thermodynamically preferred glycosidic conformation is syn, such as the 8-bromo-dG and 8-methyl-dG, if incorporated into syn dG positions generally enhance the thermodynamic stability of a GQ and generally leads to destabilization. In some cases, it also induces topological rearrangement of the GQ. In this way, all 8-substituted 2'-deoxyguanosines whose thermodynamically preferred glycosidic conformation is syn, such as the 8-bromo-dG and 8-methyl-dG, if incorporated into syn dG positions generally enhance the thermodynamic stability of the native GQ, whereas if incorporated into anti dG positions they generally reduce the stability. Similarly, if the anti stabilized sugar analogs substitute for syn dG positions, these not only destabilize the structure but can also alter the folding topology. For instance, an antiparallel fold, such as that of the monomolecular TBA GQ, transforms into a nonmonomolecular parallel fold on the effect of such substitutions. Another factor in the stabilization-destabilization is the sequence position within the potential GQ-forming oligonucleotide. The same nucleotide analog, independently of the syn/anti isomerism, can increase the stability at a given position and decrease it in another location. This alternate effect depends on the role of the substituted position in the stabilization of the folded native structure.

3.1. Incorporation of Base Derivatives

3.1.1. Purine Analogs. In GQs, the position C8 and also one hydrogen atom of NH₂ of C2 of guanine are the only suitable sites for modification that may not cause major deformation of the folding topology. Substituents of these positions point towards the grooves, towards the solvent, away from the Hoogsteen-type circular double H-bonding structure, and away from the central cavity of the stacked G-tetrad. C8 of guanine has been the most frequently modified position. The energetically preferred N-glycosidic torsion angle of 2'-deoxyguanosine (dG) is in the anti range (Figure 6, 1), whereas bulky atoms, like the halogens, the oxygen, and groups that are large enough to sterically interfere with the 5'-OH of the sugar moiety, turn the glycosidic bond into the syn-region (Figure 6, 2–6) [62–64]. The difference between these conformations has been estimated to be ~1-2 kcal/mol [33]. When the 8-substituted-dG replaces an originally syn dG, it can generally stabilize the scaffold for which the best example is 8-bromo-dG.

8-Bromo-2'-deoxyguanosine (br8dG). Dias et al. [64] described in 1994 that incorporation of br8dG into the 20-mer (T₅G₂)₄, which folded into a monomolecular antiparallel chair-type GQ (Figure 2, 2), stabilized the (T₅G₂)₄ by 5-6°C in the Tₘ value when syn dGs were switched to br8dGs. Free energy
The analog destabilized the GQ in the 2.5\degree G2 and G6 by 8.8 kcal/mol. Free energy difference between the parent and the eight syn- and anti substituted br\textsuperscript{8}dG-modified GQs was ±1 or −1 kcal/mol. Xu et al. [65] found in 2006 that br\textsuperscript{8}dG stabilized the GQ fold of the 22-mer human telomere sequence A(ttt21) or AG\textsubscript{3}(TAG\textsubscript{3})\textsubscript{5} in K\textsuperscript{+} solution when br\textsuperscript{8}dG was incorporated in place of a syn dG position of the 3-tetrad monomolecular GQ. The enhanced thermodynamic stability was supposed to be the driving force for the originally polymorphic folds [8] (Figure 2, 1–4) to convert into a more stable single form. As a result, br\textsuperscript{8}dG has since been frequently used to stabilize particular DNA and RNA GQ conformations for structural studies carried out with CD, UV, NMR spectroscopy, and other techniques [66–69]. In certain cases, br\textsuperscript{8}dG could stabilize a GQ also in anti G positions. Aviñó et al. [94] substituted the TBA GQ (Figure 2, 7) with br\textsuperscript{8}G, which stabilized the GQ when substituted for the syn position G5, by 7.3\degree C, and interestingly also by the double substitutions for G2, G11, by 3.6\degree C, which were both anti positions. The TBA GQ remained antiparallel. The analog destabilized the GQ in the anti position G2 by 2.5\degree C and also by double substitutions of the anti positions G2 and G6 by 8.8\degree C [94].

A remarkable example of using br\textsuperscript{8}dG for structure studies was the stabilization of individual topologies in telomeric GQs. Attaching flanking nucleotides to the basic htel-21 DNA sequence of G\textsubscript{3}(TTAG\textsubscript{3})\textsubscript{5} induces diverse folds [8]. For instance, TA(ttt21) was found by NMR technique to mostly fold into the hybrid-1 form (Figure 2, 3), whereas the TA(ttt21)TT adopted the hybrid-2 topology (Figure 2, 4) as the dominant, but not the only form. The hybrid structures are generally polymorphic in solution and are interconvertible due to the small energetic differences between them [84]. Incorporation of br\textsuperscript{8}dG into selected syn dG position(s) converted the polymorphic fold into a stabilized major fold. For example, the syn dG6 in the middle tetrad of the TA(ttt21) and the syn dG15 in one of the terminal tetrads of the TA(ttt21)TT GQs greatly stabilized their respective major topologies, according to Phan et al. [71]. Another noteworthy illustration for the effect of br\textsuperscript{8}dG was the induction of conformational changes in GQs. Using the same 25-mer TA(ttt21)TT, An et al. [87] incorporated two br\textsuperscript{8}dGs into the sequence. The dual incorporation induced exclusively either the hybrid-1 or the hybrid-2 fold in K\textsuperscript{+} (wt for the wild-type):

**Wild-type (wt):** TA GGG TTAGGG TTAGGG TT

br\textsuperscript{8}dG in G16, G21: TA GGG TTAGGG TTAGBr\textsuperscript{8}GG TTABr\textsuperscript{8}GGG TT hybrid-1

br\textsuperscript{8}dG in G10, G15: TA GGG TTAGBr\textsuperscript{8}GG TTABr\textsuperscript{8}GGG TTAGGG TT hybrid-2

br\textsuperscript{8}dG impacted both the topology and stability when incorporated into tetramolecular parallel GQs (Figure 2, 6). TQ\textsubscript{3}T molecules in K\textsuperscript{+} fold into [TGGGT], [TTGGGT], and TGGGGT, according to Phan et al. [67] separately substituted each G of TG\textsubscript{3}T with br\textsuperscript{8}G. All three modified sequences assembled into tetramolecular GQs; the stabilities and topologies were, however, different. CD spectrum of [TBr\textsuperscript{8}GGGT], showed, surprisingly, antiparallel characteristics and its thermal stability was much higher, by 13.3\degree C, than that of the native parallel tetramolecular scaffold. The [TGGGTT]\textsubscript{4} formed parallel GQ architecture, as the native sequence did, and stabilized the native structure by 3.6\degree C. The TGGGGGT proved unstructured, and no Tm could be determined. It was also an unusual finding that the br\textsuperscript{8}G tetrads were all syn with each tetramolecular analog, while the other two G-tetrads of the GQ were anti, as in the case of the unmodified [TGGGT].

8-Methyl-2'-deoxyguanosine (m\textsuperscript{8}dG). Another stabilizing guanine analog is the 8-methylguanine nucleoside (m\textsuperscript{8}dG) (Figure 6, 3), and as with the br\textsuperscript{8}dG, it is stabilizing only if it is substituted for syn dGs. m\textsuperscript{8}dG has been, however, much less utilized than the br\textsuperscript{8}G. He et al. [33] incorporated a methyl (and a 1-propynyl) group into the C8 positions of G1, G5, G10, and G14 of the 15-mer TBA DNA oligonucleotide (Figure 2, 7), where positions were all syn. Thermal stability of the unmodified TBA GQs, 55\degree C, was increased to 70\degree C by these substitutions. Concomitantly, the substitutions increased the thrombin-binding activity as well, presumably by the stabilization of the chair-like structure. Xu and Sugiyama [95] incorporated the m\textsuperscript{8}dG into the 18-mer CGGGGGGTTTTGCGCGGC of the G-rich termini of retinoblastoma gene DNA. Thermal stability of the two-tetrad antiparallel GQ increased when syn dGs were replaced and decreased when an anti dG was substituted by m\textsuperscript{8}dG. The GQ stability pays energy penalty for accommodating a syn dG analog in place of an anti dG. Later, m\textsuperscript{8}dG was incorporated into the tetramolecular GQ-forming TG\textsubscript{3}T and TG\textsubscript{3}T oligodeoxynucleotides (Figure 2, 6). Using [TGGGT], Virgilio et al. [96, 97] and Tran et al. [98] determined that m\textsuperscript{8}dG had a position-dependent effect on folding arrangements. N-glycosidic linkages of the G and T nucleosides are all anti in the native fourfold parallel GQs. In position 2, however, m\textsuperscript{8}dG formed an all-syn tetrad that resulted in an antiparallel-type CD spectrum in K\textsuperscript{+}, that is, a large positive band near 290 nm and a negative one near 260 nm (Figure 3). The m\textsuperscript{8}dG-tetrad in position 3 was all anti, the GQ was parallel, and the CD spectrum showed large positive peak at 260 nm and a small negative one around 240 nm. m\textsuperscript{8}dG in position 4 hindered the GQ formation. Results were similar to those observed with br\textsuperscript{8}dG. The authors also studied m\textsuperscript{8}dG in position 2 of the anti-HIV aptamer TGGGGAGT and observed thermal stabilization. Again, all-syn m\textsuperscript{8}dG tetrads were observed in both GQs; however, the topologies remained parallel [99]. Structure and kinetics of formation of the tetramolecular GQ model were investigated using m\textsuperscript{8}G in another study, m\textsuperscript{8}dT at the 5'-end of the short oligonucleotide accelerated GQ formation by 15-fold relative to the unmodified oligonucleotide [98]. In a recent study, Zhao et al. [100] substituted dGs by m\textsuperscript{8}dGs in the GQs
formed by an htel, the TBA, and the c-myc promoter's G-rich pu27 sequences. m²dG proved significantly stabilizing in those cases when it replaced syn dGs and the stabilization was cumulative when two or three m³dGs were substituted for syn dGs. However, m⁶dG proved destabilizing when an anti dG was replaced.

8-O-Methyl-2'-deoxyguanosine (om⁸dG). Lech and coworkers [80] studied the influence of 8-O-methylguanine deoxyguanosine (Figure 6, 4) on GQ structures using the 24-mer model TT(htel-21)A that forms a predominant hybrid-1 topology (Figure 2, 3) in K⁺ solution [101]. Inserted one by one into six different positions (bold) into the 24-mer TTGGGTTAGGGTTAGGGTTAGGA, each analog elevated the thermal stability of the GQ by an average of 3.1 ± 1.4°C, with slight positional effects. Except for position 22, each site was syn. The guanine analog in the anti dG22 position also elevated the stability, although somewhat less than in the 21 syn position. The hybrid-1 topology was not basically changed by om³dG.

8-Amino-2'-deoxyguanosine (n⁸dG). The effect of the syn-prefering n⁸dG on GQ stability (Figure 6, 5) was first described using the TBA GQ by De La Osa et al. [102]. The analog was placed (for MD calculation) and incorporated into position 2, which is an anti dG position of the GQ. Molecular dynamics and thermodynamic integration calculations (MD/TI) suggested that the n⁸dG substitution led to destabilization by 1.2–1.9 kcal/mol in the ΔΔGᵣ value, and the nature of the central cation (Na⁺ or K⁺) did not have substantial effect. CD-based melting profiles confirmed the MD/TI calculations. This destabilization was enthalpic in nature, with a ΔΔHᵣ of 5.5 kcal/mol. n⁸dG was later incorporated into the tetramolecular parallel GQ models built from TGᵣT and TGᵣT and an unusual position-dependent effect was observed by Gros et al. [73]. In the [TGᵣT]₄, the n⁸dG tetrads reduced the thermostability in three of the four G positions and the stability only in position 3 of the TGᵣT sequence. With another, on the 5'-end covalently linked tetramolecular assembly, Ferreira and coworkers [103] found that inserting a single n⁸G only into one strand in position 2 or 3 stabilized the parallel GQ. In both cases, n⁸G-G-G-G tetrads were formed:

\[ \text{3'}-\text{Tn}^8\text{GGGGTT}-5'; \text{TB}-5'-\text{TTGGGGTT-3'} \]
\[ \text{3'}-\text{Tcn}^8\text{GGGGTT}-5'; \text{TB}-5'-\text{TTGGGGTT-3'} \]

where TB was the Trebler linker [O-phosphate-CH₂-C(CH₃-OCH₃)CH₂-O phosphate]. With the intramolecular, mostly hybrid-1 GQ of TT(htel-21)A, Lech et al. [80] found that n⁸dG was only marginally stabilizing when incorporated into the syn dG positions of 3, 9, 15, 16, and 21 with the average ΔTₘ of 1.1 ± 1.1°C. According to the CD spectra the hybrid-1 type, topology (Figure 2, 3) apparently was not changed by any of these substitutions.

8-Propynyl Derivatives of 2'-Deoxyguanosine. The triple C-C bond-containing propynyl group in conjugation with heteroaromatic bases, such as the uracil in duplexes [104] as well as in the TBA GQs [52], greatly enhanced the stability of the macromolecular structure through strengthening the stacking. The propynyl group in conjugation with guanine (Figure 6, 6) also stabilized the GQ, which was again the TBA GQ, when incorporated into all four syn dG positions, G1, G5, G10, and G14, of the 15-mer sequence. The substitutions increased the thrombin-binding activity of modified TBA GQ, as compared to the wild type, probably due to the stabilization of the intramolecular GQ structure. Larger substituent groups in these positions, like phenyl-ethynyl, however, decreased the activity, probably due to steric hindrance [33].

8-Vinyl-2'-deoxyguanosine (v⁸dG) and Its Derivatives. The fluorescent base 8-vinylguanine (v⁸G) (Figure 7, 1) proved to be a good alternative to the widely used fluorescent marker 2-aminopurine (n²Pu, Figure 8, 8) [105, 106]. Nadler et al. [105] incorporated the v⁸dG into positions G3, G15 (middle tetrad), and G4 (terminal tetrad) of the htel-23 A(htel-21)T. In 135 mM Na⁺ solution, the two middle tetrad substitutions elevated the thermal stability of the wild type (Tₘ(¼) = 56°C) by 3-4°C and v⁸G in one of the two terminal tetrads by 2°C. In 100 mM K⁺ only the vinyl analog of the middle tetrad (position 15) increased the stability by 5°C of the unmodified GQ (Tₘ(¼) = 61°C). CD spectra of the three modified GQs in Na⁺ looked hybrid types contrary to the wild-type's antiparallel spectrum (Figure 3). The wild-type htel-23 folds into hybrid conformations in K⁺ solution. The middle tetrad-modified GQ studied followed the trend in K⁺, whereas the CD spectrum of the terminal tetrad-modified GQ was antiparallel type [105]. 8-Vinylguanine was also attached to 2-aminoethylglycine, creating a peptide nucleic acid (PNA) building block. This allowed the differentiation between topologies of GQs based on fluorescence intensity. v⁸dG was found to be capable of adopting both syn and anti conformations required by distinct GQ structures [106].

Aromatic derivatives of 8-vinylguanine, the 8-(2-phenyl-ethynyl)- and 8-[2-(pyrid-4-yl)-ethynyl]guanineines (Figure 7, 2, 3), have been studied by Dumas and Luedtke [107] by incorporating them into positions 9 and 23 of the 24-mer TT(htel-21)A. Although v⁸G can adopt both syn and anti conformations [105, 106], the two aromatic derivatives of v⁸G adopted only anti glycosidic conformation in solutions, like dG does, contrary to other 8-modified dGs, such as br³-, m⁸-, n⁸-, or o⁰dG. Due to the conjugated vinyl tethers, the bulky aryl and heteroaryl groups are detached far enough from dG not to hinder the energetically preferred anti orientation of dG. The GQ of TT(htel-21)A has been described to fold into hybrid-1 topology (Figure 2, 3) in K⁺ and into the antiparallel topology in Na⁺ (Figure 2, 1) [107]. The position 9-substituted 24-mer GQ was stabilized by the two ethynyl derivatives by 7-8°C in Na⁺ and 5-8°C in K⁺; however, in position 23, the GQs were destabilized by 1-6°C in Na⁺ and 3-8°C in K⁺. In position 9, the two G-analogs did not change the CD spectrum of the wild-type GQs.
Figure 7: Structure of the syn 8-vinyl-2-deoxyguanosine (1); the anti 8-(2-phenylethenyl)- (2) and 8-[2-(pyrid-4-yl)-ethenyl]-dGs (3); the cis (4) and trans isomers (5) of the anti 8-vinyl-substituted-dGs; and the syn 8-(2-pyridyl)-dG (6).

(hybrid), whereas the position 23-substituted GQs showed antiparallel-type spectra in K⁺. In Na⁺, the modified GQs remained antiparallel like the wild-type tetr-24 GQ did [107]. 8-Fluoroenylvinylguanine (Fv⁸G), another derivative of v⁸G, was also incorporated into GQ-forming oligonucleotides. Ogasawara and Maeda [108] developed a light-controlled reversible folding-unfolding GQ structure with this analog based on the cis-trans isomerization of a photochromic nucleobase Fv⁸G (Figure 7, 4, 5), incorporated into the 15-mer TBA sequences, the GGTTGFv⁸GTGTGv⁸G TTGG and Gv⁸GFv⁸G TTGGTGTGTTGFv⁸G. The native TBA sequence folds into a chair-type, two-tetrad antiparallel GQ in K⁺ (Figure 2), and its CD spectrum was only slightly changed by the trans-Fv⁸G (Figure 7, 4) in anti dG positions, G6 and G11 and G2 and G15. Tₘ of the modified GQs increased by ∼10°C, up to ∼60°C. On irradiation at 410 nm, the trans-Fv⁸G changed into the cis form (Figure 7, 5), which resulted in unfolding of the stable GQs. Irradiation at 310 nm reversed the unfolding. 8-(2-Pyridyl)guanine (Py⁸G, Figure 7, 6) is a highly fluorescent compound that has been used to study the folding of the 24-mer TTG₃(TTAG₃)₃A GQs [109]. Positions G9 and G23 of terminal tetrads were syn in Na⁺; still the syn Py⁸dG affected the thermostability in altered ways: 9-Py⁸dG increased Tₘ of the wild-type GQ by 7 and 10°C in Na⁺ and K⁺, respectively, whereas 23-Py⁸dG reduced the wild-type's Tₘ by 2 and 9°C, respectively. Apparently, the glycosidic orientation of the Py⁸dG was not related to these effects. 9-Py⁸G in K⁺ did not change the wild-type’s hybrid fold; however, 23-Py⁸G induced a conversion into antiparallel topology in K⁺.

The “push-pull” emissive fluorophores, which can exhibit environmentally sensitive quantum yields due to excited-state proton-transfer reactions with bulk solvent, were thoroughly investigated by Manderville’s and Wetmore’s groups. 8-Furyl- and 8-(4-cyanophenyl)-2'-deoxyguanosines (Figure 8, 1, 2) were used to monitor the duplex-GQ exchanges. In a syn dG position of the TBA GQ (pos. 5), the syn 8-aryl dGs significantly stabilized the GQ in both Na⁺ and K⁺ solutions, whereas they substantially destabilized it when incorporated into an anti dG location (pos. 6) [110, 111].
8-furyl- and 8-vinyl-benzo(b)thienyl-dGs (Figure 8, 1, 3), which, based on analogy, are assumed to be syn and anti dGs, respectively, were incorporated into both syn and anti dG positions of the TBA DNA. 8-Furyl-dG in the syn positions of 10 and 14 elevated $T_m$ of TBA GQ, while the more lipophilic benzothienyl derivative decreased it slightly in the syn G5 position and more extensively in the anti G6 position. Double labeling with furyl and benzothienyl derivatives enhanced the destabilization when one of the analogs was in anti position; however, when both probes substituted for syn dG positions, the wild type was significantly stabilized [112]. The push-pull phenomenon in the TBA GQ was also studied with acetylphenyl, benzo[b]thienyl, quinolyl, pyren-1-yl (Figure 8, 7), and the 8-vinyl tethered derivatives of acetylphenyl and benzo[b]thienyl attached to C8 of dG (Figure 8). The dG analogs were substituted for syn dGs of the TBA DNA sequence. The aromatic and heteroaromatic groups with the aryl ring attached directly to C8 of guanine thermodynamically favor the syn glycosidic conformations and when these analogs substituted for the syn dG5 they increased the thermostability of the wild-type TBA GQ. The two aromatic analogs tethered to guanine base via the conjugated vinyl group favored anti glycosidic torsion angle (Figure 8, 3, 5). The anti 8-vinyl-acetylphenyl-dG slightly decreased $T_m$ of the wild type in the syn dG5 position but moderately elevated it in the anti dG6 position [113]. Similarly, the fluorescent 8-pyrrolyl-, 8-furyl-, 8-thienyl-, 8-benzofuryl-, 8-indolyl-, and 8-benzothienyl-dGs were also incorporated into the TBA sequence in another study. These nucleosides also preferentially adopt syn conformation in solution and their insertion into the syn dG5 enhanced the thermal stability of the parent GQ in K+ solution by 1–11°C and did not perturb the thrombin-binding affinity. 8-Thienyl-dG was found to increase the thermostability in each of the four syn positions by 7–10°C and its double incorporation by 18.5°C, and substituting all four syn dGs, $\Delta T_m$ was higher than 36°C [114]. Stability changes resulting from the syn versus anti glycosidic conformations of dG derivatives were similar.
to 8-(p-cyanophenyl)-dG in TBA and 8-furyl-dG in A(htel-21) GQ systems [115]. 8-Furylguanine was also incorporated into each of the three tetrads of A(htel-21) and the effects of salts and cosolvents (acetoni-trile, PEG-600, and N-methylmesoporphyrin) were examined on the structure. In addition to the influence of the syn and anti positions on stability, the position of the tetrad substituted affected the stability too [116]. The effect of 8-vinyl-acetylphenyl- and 8-vinylbenzothienyl-dGs on the TBA GQ system was also investigated for the detection of divalent metal ions. Based on its emission sensitivity to Pb²⁺, the vinyl-benzothienyl derivative proved to be an effective and selective sensor for Pb²⁺, having higher affinity for Pb²⁺ than for Na⁺, K⁺, or other bivalent cations of biological importance [117].

The C8-aryl- and heteroaryl substituents, such as C8-phenyl, pyridine, thiophene, and furan of dG, were also studied by Dumas and Luedtke [118], who used the changes in fluorescence to characterize the metal-binding affinity and specificity of the 8-substituted guanines in duplex and GQ DNAs as well as to study the effect on stability. Thermal stability of the GQ formed by the C8-substitution AG₈AG₉CXCTG₃XAG₉AG₉ substituted in positions 10 and 15 (X) was investigated. 8-(2-Pyridinyl)guanine (Figure 7, 6) decreased Tₘ of unmodified GQ (64.6°C) by 2–4°C in positions 10 and 15. This analog has also been used to study the folding properties of the 24-mer T'T(htel-21)A GQ [109]. Position G23 in the 3'-terminal tetrad was syn in Na⁺; still the syn-Py²⁺dG23 reduced the wild-type’s Tₘ by 2 and 9°C in Na⁺ and K⁺, respectively. Apparently, the glycosidic orientation of the Py²⁺dG was not related to the effect. The Py²⁺G23 induced a conversion of the hybrid fold(s) into antiparallel topology in K⁺.

2-Aminopurine (n²Pu). The fluorescent 2-aminopurine (n²Pu, a G or A analog) (Figure 8, 8) incorporated into the loop sequence of a potentially GQ-forming oligodeoxynucleotide has significantly enhanced fluorescent emission upon the formation of the GQ. Its fluorescence is quenched in duplex DNA due to the stacking with flanking bases and becomes comparable to that of the free n²Pu upon the formation of GQ as the π-stacking within the loops is distorted and effective quenching cannot occur.) n²Pu has been widely used in structural analysis of GQs. For example, for conformational studies, n²Pu was substituted separately for all four adenines in A(htel-21) [84, 119–121]. The analog did not change the folding and only slightly reduced the thermal stability of the parent GQ [122], while significant changes were observed in fluorescence intensity of n²Pu depending on whether it was present in the GQ or in the duplex formed with the complementary strand. n²Pu could also distinguish between the basket-type and propeller-type GQs. 2-Aminopurine has also been used to develop a sensitive fluorescent GQ assay for uracil DNA-glycosylase activity [123]. 2-Aminopurine was also substituted for Ts of the loops the 15-mer TBA and the (G₇T₈)₃G₉ oligonucleotides to develop fluorescent detection methods. The two oligomers folded into antiparallel and parallel GQs, respectively, in K⁺. The substitutions did not change the original topologies but changed the stabilities of the wild-type Qs. Tₘ of 48°C for wild-type TBA GQ in 50 mM KCl was increased by 2°C when substituting for T3 and T2 of the loops, in which positions the T is stacked to the core [18–20]. In the other T positions, the n²Pu reduced the stability to around 40°C. Tₘ of the (G₇T₈)₃G₉ GQ measured in 10 mM KCl was 90°C and was reduced in all three loop T positions by n²Pu. Single substitutions caused 4.5°C decrease in Tₘ, while double and triple substitutions caused a decrease of 12°C and 16°C, respectively [124]. Similarly, n²Pu substituting for G8 of the central TGT loop just slightly reduced the Tₘ value of 52.4°C of the native GQ to 51.6°C, but the modified TBA GQ proved more stable based on the free energy change (ΔG⁻²₀) value from −2.3 to −3.6 kcal/mol [125].

5-Nitroindole. Tsvetkov et al. [126] incorporated the universal base 5-nitroindole (Figure 8, 9) into various positions of the TBA oligonucleotide. All modified TBA forms antiparallel GQ structures and retained the ability to inhibit thrombin. UV absorption-based thermal stability of T7 substitution resulted in Tₘ of 43.0°C and of the T9 in 38.5°C, while Tₘ of the native TBA GQ was 51.9°C. On the other hand, replacement by 5-nitroindole of the anti G8 retained the stability of the native TBA GQ and this substitution substantially increased the clotting time and resulted in a twofold lower IC₅₀ value, as compared to the unmodified TBA GQ. Attachment of 5-nitroindole either to the 5’ or to the 3’-terminus did not change the stability of the unmodified GQ, 51.1 and 51.2°C, respectively [126].

3.1.2. Pyrimidine Derivatives. Loop length and composition are the major factors that determine the stability and folding features of natural GQs [61, 127, 128]. For instance, a single one-nucleotide loop generally keeps the GQ in parallel topology [129, 130]. With increasing loop lengths, thermodynamic stability generally decreases [128]. Effect of loop substitutions on GQ structures has been widely studied, but the majority of the investigations focused on the mutation by natural nucleotides, T/U, C, A, and I, to increase the loop length and change the sequence. The TBA DNA was among the most mutated GQs due to its stable and straightforward conformation [10]. In addition to 2-aminopurine, examined in the preceding section, those few loop base modifications are discussed here that have been carried out using the synthetic nucleotide analogs only, which increase or decrease the stability depending on the position of substitution. The natural base lesions that also form in the loops of GQs, such as o₃A, hm²U, and AP sites, are reviewed in Section 5.

The TBA DNA’s six loop thymidines were replaced by a fluorescent furyl derivative, the 5-furyl-2’-deoxyuridine (Figure 8, 10), to demonstrate the analog’s ability to determine the positional impact of each thymine on the stability and thrombin-binding activity of the TBA GQ [131]. Earlier NMR studies suggested that T4 and T13 stack strongly with the neighboring G-tetrad and T9 of GGTGGTGGTGG interacts with G8 of the TG₃ loop and with the G-tetrad. The T3, T7, and T12 do not interact with adjacent nucleotides and point outwards, towards the solvent [18–20]. Thermal stability of the modified TBA GQs showed good correlation with the steric predictions of the NMR studies. In K⁺ solution,
Figure 9: Structures of the anti ribo- (1) and the 2’-O-methylribouridines (2); the C2’-endo (S-type) and C3’-endo (N-type) sugar puckers (3); the 2’-fluoro-2’-deoxyribofuranosylguanine (4); and 2’-fluoro-2’-deoxy-arabinofuranosylguanine (5).
were observed when the gggg quartets were at the 5'-end of the G stacks in both systems. However, gggg quartets replacing the other dG quartets stabilized the GQ structures.

A single 2'-OMe modification significantly enhanced the thermal stability, by close to 10°C, and also the peroxidase activity of the parallel GQ-forming sequence [B7]-3-0 [142]. Single uridine units were inserted into the three loops of the TBA GQ, separately. The uridine replacing thymidines in the central TGT loop resulted in an increased stability of the antiparallel GQ; however, substitution of thymidines in the TT loops induced destabilization of the TBA GQ [143].

2'-Fluoro-2'-deoxyribonanosyl- and 2'-Fluoro-2'-deoxyarabino- 

binosyl Nucleosides. Several laboratories have investigated the effect of 2'-deoxy-2'-fluoro-D-arabinonucleotides on GQ structures (fl araN), where N is any base (Figure 9, 5). The antiparallel TBA, the bimolecular [G4T4G4], and the tetramolecular phosphorothioate backbone-containing anti-HIV [T2G1T2] were among the first GQ models examined [144]. The energetically preferred N-glycosidic conformation of the fl araN is anti. Incorporation of fl araG or fl araT in place of anti dG-s of the tetrads or Ts of the loops in the antiparallel or parallel GQs stabilized the structure and maintained the fold of the wild-type GQ [144]. Lech's group [145] also investigated the effect of fl araG and the 2'-fluoro-2'-deoxyguanosine, fl dG (Figure 9, 4). The deoxy-GQ models were the hybrid-1 TA(hotel-21) [101] and the hybrid-2 TA(hotel-21)/TT in K' [146]. Substitution of the anti dG in the middle tetrad by either of the two nucleoside analogs increased the dominance of single folds and also enhanced the thermal stability by about 3°C. However, when all five syn dGs were replaced by the analogs, the hybrid forms switched to parallel topology and the thermal stabilities increased by about 10°C, as compared to the stability of the respective wild type. Single riboguanosine (g), arabinofuranosyl-guanine (araG), and 2'-deoxy-2'-fluoro-arabino-furanosylguanine (fl4araG) replaced the dG9 of the 12-mer TAGGGTTAGGGT [147] that builds a mixture of two interconverting folds in K' solution: a dimeric parallel and a dimeric antiparallel GQ [66]. The dG9 was syn in the antiparallel form; thus its replacement by the anti favoring sugar analogs converted the mixture to the dimeric parallel GQ with the concomitant increase in thermal stability from 41°C to 47–53°C. The fl dG furnished the greatest stabilization, 12°C [145]. The effect of a single 2'-fluoro-2'-deoxyguanosine (fl dG), 2'-deoxy-2'-fluoro-arabinofuranosylguanine (fl4araG), and 2'-O-4'-C-methylene-guanosine (LNA) (see next section and Figure 10, I) was analyzed with more than 60 parallel and hybrid GQs [148]. Generally, substitutions of anti dGs of the G-tetrads increased the stability of GQs, while substitutions of syn positions disrupted the native GQ conformation. The 22-merGGATGGGACACAGGGACGGG oligonucleotide forms into a unimolecular hybrid-type GQ in K' solution. 2'-Fluoro-2'-deoxyguanosine (fl4 dG), which also favors the anti glycosidic conformation, was substituted for the syn GI, G6, and G20 (bold, underlined) of the 5'-terminal tetrad. The substitutions changed the polarity of the tetrad. The overall fold, however, did not change and thus created a novel type of GQ in which all four G-columns comprised only syn or anti dGs: one column with all-syn and three G-columns with all anti glycosidic linkages [149]. Dickerhoff's group using fl dG with the GQs of the previously applied 22-mer G3ATG3ACACAG3GACG3 and now also with the TT(hotel-21)A confirmed the fl-dG-induced conformational perturbations [150]. A single fl dU was also inserted into the three loops of the TBA GQ. Substitution of thymidines in the TT loops by fl dU resulted in destabilization of the TBA GQ [143].

Lietard et al. [151] synthesized the first microarrays containing fl araN and fl araN/DNA chimeric TBA oligonucleotides to fully map the binding affinity landscape of the TBA GQ. A series of promising fl ara-modified aptamer candidates were found with Kd values that are significantly lower than that of the unmodified TBA GQ and which adopted highly stable, antiparallel GQ structures. The presence of fl araT at position T3 not only drastically strengthened the binding but also elevated the thermal stability of the unmodified TBA GQ by 7°C (Tm; 47°C). The TBA GQ could accommodate up to ten fluoro-ara nucleotides, in which TBA analogs were also promising aptamer candidates with dissociation constants up to three times lower than that of the wild-type TBA GQ. With these analogs' considerable increase, up to 20°C of GQ stability was observed and stable folding was observed even in the absence of K+ ions. The T3 modification apparently preorganized the dinucleotide loop into the proper conformation for interaction with thrombin.

Locked Ribonucleosides (LNA). LNA is the abbreviation for locked nucleic acid. The LNA nucleoside has a modified ribose ring, which is locked in a stable C3'-endo or N-type (RNA-like) sugar puckering by a 2'-O-4'-C-methylene linkage. Thus, the LNA contains a bicyclo-sugar moiety, resulting in a stable anti N-glycosidic conformation of the nucleoside (Figure 10, I). The LNA nucleosides increase the stability of duplex and triplex DNAs [152] but the B-type duplex is converted to the RNA-like A-type geometry [153]. LNA oligonucleotides have excellent hybridization properties with DNA and RNA oligomers. The structural impact of LNA in parallel GQs has been characterized by several groups [154, 155]. Nielsen et al. [156] used the telomeric sequence from the Oxytricha nova, G4T4G4 for a GQ model. In K' this oligonucleotide forms a dimeric GQ with antiparallel G-columns and diagonal T4 loops [157, 158]. The LNA-substituted sequence GLGLT, GLGL, where L stands for LNA-G, also formed a dimeric GQ, in which each G-stretch folded back into a V-shaped turn [159] and interacted with each of the three other G-stretches through formation of four G-tetrads. This new GQ folding topology was named the V4 fold. The V4 fold incorporates the features of both parallel and antiparallel GQs in one structure and the remarkable folding topology leads to tetrad steps with both coaligned (between outer and inner G-tetrads) and antialigned (between the two inner G-tetrads) hydrogen bonding, as in the parallel and antiparallel GQs, respectively. CD spectrum of the V4 fold thus displayed signatures of both parallel and antiparallel stacking [156]. In the intramolecular antiparallel, basket-type GQ-forming (G4T4G4), a single G-LNA or T-LNA reduced the stability of
the native GQ in a position-dependent way, most extensively by the G-LNAs [160]. A single G-LNA incorporated at the 3′-terminal G-position of TBA GQ did not change the topology but destabilized the wild-type GQ, by 6°C, from $T_m$ of 52°C to 46°C [161]. The TBA GQ was also the model for G-LNA and T-LNA substitutions at G2 (anti), G5 (syn, both in the same G-tetrad), and T4 (in a TT loop), T7, and G8 (anti, in the TGT loop) positions. G-LNA at position G2 reduced $T_m$ of the unmodified GQ to the largest extent, from 48.1 to 33.5°C in 50 mM KCl, although dG2 is anti. Substitution of T7 also reduced $T_m$ by about 5°C, whereas G5-LNA (despite being a syn dG position) and G8-LNA (anti) increased the thermal stability by 2.6 and 1.9°C, respectively. The T4-LNA GQ was unstable, and hence $T_m$ could not be determined. Thermal stability of the LNA-GQs was more or less inversely related to their antithrombin activity. The substitutions did not change the folding topology of the wild type as Bonifacio et al. reported [162]. The full LNA analog of the TBA DNA did not fold into a GQ, probably due to the rigid nature of the loop position LNA-Ts. The LNA/DNA chimera, in which the tetrad dGs have been substituted by LNA-G nucleotides, did fold into GQ; the CD spectra reflected parallel folding with a peak max. near 260 nm. The hysteresis observed in the melting process referred to a nonunimolecular parallel scaffold. The anti LNA-Gs that substituted for syn dGs might have induced the formation of the parallel scaffold. The structure was very stable, and the $T_m$ value increased by 20°C relative to the wild type [154].

The htel-12 TAGGGTTAGGGT forms a parallel-antiparallel mixture of dimeric GQs; the latter antiparallel one contains both anti and syn nucleosides. Substitutions of the natural nucleotides by LNAs, which are restricted to the anti form, converted the antiparallel folds into parallel, which contains only anti nucleosides. The driving force for the conversion of topology is described as a combination of the C3′-endo puckering (Figure 9, 3) of the LNA nucleotides and their preference for the anti glycosidic conformation. In addition, the LNA-modified parallel GQs are significantly stabilized, by up to 11°C in their $T_m$ value, relative to their DNA counterparts [163]. LNA-G substituting for anti dG

---

**Figure 10:** Structures of the locked nucleoside 2′-O-4′-C-methylene-guanosine (LNA), where B stands for a nucleobase (1); the unlocked nucleoside 2′-3′-acyclic-ribonucleotide (UNA) (2); 4-thio-unlocked-uridine (3); 2′-C-piperazino-unlocked-uridine (4); an acyclic thymine derivative, (S)-GNA-T (5); and the beta- and alpha-thymidines (6).
positions in the parallel-stranded GQ of T₃G₃TG₃TG₃TG₃T and the hybrid-I-forming TT(htel-21)A strengthened the GQ, while substituting for syn DG positions reduced $T_m$ of the wild type or even disrupted the native GQ structure. Particularly large negative $\Delta T_m$ values were described for the parallel-forming GQ, $-41$ to $-45^\circ$C, from the $77^\circ$C of the native [148].

The tetramolecular parallel GQ [TG₄T₄]₄ was stabilized by LNAs as much as by the ribo substitution [154]. The full LNA GQ showed the smallest hysteresis on reannealing, probably due to the rigid sugar part. The elevated stability was explained by entropy gains. An NMR study of two LNA-modified [TG₄T₄]₄ GQs showed only local structural alterations, which were due to the C3′-endo sugar pucker [155]. The crystal structure of an all-LNA-substituted tetramolecular parallel GQ formed from TG₄T was determined for the first time by Russo et al., refined at 1.7 Å resolutions. A T-tetrad was formed at the 3′-end [164].

AS4011 is a GQ-forming aptamer capable of selectively entering cancer cells by nucleolin receptor-mediated uptake. Its internalization efficiency highly depends on the chemistry of the oligonucleotides. LNA substitutions were investigated entering cancer cells by nucleolin receptor-mediated uptake. [166]. A series of interesting 3′-NHCH$_2$∼Cf or the unmodified GQ was raised to $-52^\circ$C and $-46^\circ$C, respectively, in the two modified structures studied [166]. A series of interesting 3′-end-modified (capped) LNA analogs have been prepared by Kasahara et al. [167], in which the 2′,4′-methylene bridge was replaced by -CH$_2$OCH$_2$-, -NHCH$_2$-, and -CH(Ph)OCH$_2$- groups. The bridged nucleosides increased the resistance against nucleases in human serum of the TBA analogs, and the binding abilities were not affected by these modifications. Thermal stabilities were not published. Hotoda’s anti-HIV-I oligonucleotide, the tetramolecular parallel GQ-forming TGGGGAG’s thermal stability, was significantly increased by LNA modifications. The unmodified had a $T_{1/2}$ value of 55°C while $T_{1/2}$ for TGGGGAG was 75°C, where the gold-underlined Gs were LNA-Gs. The LNA modification highly enhanced the anti-HIV-I activity of the Hotoda GQ [168]. The U nucleosides of the natural RNA sequence (uuaggg), forming the TERRA GQ were also replaced by LNA and 2′-O-methyl ribonucleoside analogs to study the protein recognition of the loops ribose moieties. Stability data have not been published [169].

**Unlocked Ribonucleosides (UNA).** The UNA abbreviation stands for unlocked nucleic acid. In the unlocked nucleosides, the furanose ring’s C2′-C3′ bond is missing; it is thus a 2′-3′-acyclic-rN, a ribonucleoside analog (Figure 10, 2), which is often abbreviated as, for example, uG or uU. These types of nucleosides are flexible ribo derivatives. The positional effects of single unlocked nucleosides, the 2′-3′-acyclic-rG and 1′-U, were described with TBA GQs. Replacement of T by uU in loop positions 3, 7, and 12, one in each of the three loops, proved stabilizing for the TBA GQ. uU in the other three T positions destabilized the GQ [170]. Uracil in the loops has been described to stabilize the TBA GQ, contrary to the duplexes, by $-0.16$ kcal/mol each [1]; thus, with the uU nucleosides, this effect must have contributed to the stabilization. CD spectra of GQs with uU in positions 3, 7, and 12 were the same as that of the unmodified TBA GQ. uU in the other loop positions changed the spectra by significantly lowering the amplitudes and causing band shifts. Incorporation of uG in any tetrad position either destabilized the GQ or hindered the formation of it. The GQ with a uU in position 7 proved to be the only one with more effectiveness in blood clotting than the unmodified GQ. The influence of a single uU and 2′-C-piperazino-uU residues (Figure 10, 4) incorporated into several positions of the TBA DNA was studied by Jensen et al. [171], who arrived at very similar conclusions as did Pasternak et al. [170]. The 2′-C-piperazino-uU more efficiently stabilized the GQ structure than the uU and increased the thermal stability of the native TBA GQ by 2–3°C in a position-dependent manner. GQ topology and molecularity were retained. The presence of uU in positions U3, U7, and U12 resulted in the highest stabilization of the GQ. On the contrary, the largest destabilization mounting to $-15^\circ$C was observed when uU residues were placed in positions U7, G8, and U9. Kinetic studies indicated no strict correlation between thermodynamic stability and the binding affinity to thrombin. Most variants studied bound to thrombin, albeit with decreased affinity related to the wild-type TBA GQ [171]. A double modification study of the TBA with an UNA analog, the unlocked 4-thiouridine, the sUuU (Figure 10, 3), has also been published. The analog in all possible positions of the three edgewise loops produced negative Gibbs’ free energy indicating that at physiological temperature the predominant species were the folded GQ forms. However, most TBA variants were less stable than the unmodified GQ by 0.9–1.0 kcal/mol. sUuU at positions 3 and 12 did not influence the thermodynamic stability, whereas, at position 7, it increased it by 0.34 kcal/mol. In contrast, the ribo 4-thiouridine introduced into positions 3, 7, 9, and 12 stabilized TBA GQ by 0.31–0.53 kcal/mol and was usually more stabilizing/less destabilizing than the sUuU. The modified TBA retained the antiparallel GQ topology of the native GQ. Thrombin clotting time studies revealed that TBA modified with sUuU at position 7 possessed high anticoagulant activities and the modified aptamer was a potent inhibitor of fibrin-clot formation [172]. Unlocked uA, uT, or uC was substituted for the single-base diagonal loop and uG for tetrad Gs in the GQ of G₃T₃G₃A(or T,C)G₃G₃T₃G₃. The loop (A, T, or C) modifications stabilized the GQ by 3–7°C in $T_m$, and the uA had the greatest effect. Stabilization was explained by the flexibility of the unlocked nucleoside, which could ease the tension that might exist in a single-base diagonal loop. Contrary to the loop substitutions, the uGs led to significant destabilization of the GQ. The uG in the middle tetrad caused the largest effect, and $\Delta T_m$ was $-176^\circ$C. Flexibility of the sugar moiety did not prove to be a beneficial structural motif in the G-tetras. The unlocked nucleosides caused transition from the antiparallel to hybrid-type fold in some of these GQ analogs [173]. Thymine glycol, also an acyclic analog of dT, (S)-GNA-T (Figure 10, 5), was substituted for loop T position of the TBA GQ. The analog destabilized the TBA in positions in T4, T9 by 9–13°C, and stabilized it in T7 by 5.2°C. With double substitutions in T7,9, T3,12,
and T4,T13, the TBA GQ was destabilized by 2–12°C [94]. Aldering et al. [57] reported on the impact of uU and the 3′-amino-uU on the structural dynamics and stability of TBA GQ by substituting them for the (nonstacking) loops T3, T7, and T12. While the uU enhanced the thermal stability of the unmodified GQ (TmC 49°C) at all three T positions, by 2–5°C, the 3′-amino derivative decreased it by 2–7°C. As the amino group replaced the 3′-OH, the internucleotide linkage at this site was 2′-5′ instead of the natural 3′-5′ (this modification of the backbone is named isoDNA; see also later). The authors suggested that the altered backbone caused the destabilization effect with the 3′-amino modifications. On the other hand, none of the loop substitutions changed the topology of the native TBA GQ (CD peak max. at 295 nm and negative peak at 265 nm), and both the unlocked uridine analogs improved the thrombin clotting time [57].

α-Anomeric Sugars. α-Nucleosides differ from the natural β-ones by the inversion of the configuration at the C(1′) anomeric position of the furanose ring (Figure 10, 6) (for a review, see [174]). The T3,G4,T5 that assembles into a parallel-stranded tetramolecular GQ was found to be an inhibitor of HIV infection, especially if the phosphodiester internucleotide linkages were replaced with P-S linkages (see Figure 3.3), and its IC50 value was 0.3 µM [175]. When the same oligodeoxynucleotide was built from α-nucleosides, the α(T3,G4,T5), also with P-S linkages, a similar anti-HIV activity was observed, and the IC50 value was 0.5 µM [176]. Since α-oligonucleotides are nuclease-resistant [174], it was claimed that the P-S backbone is mechanistically required for antiviral activity of this oligonucleotide. It was, therefore, suggested that PS-ODN interacts with the highly cationic topositoft he nat iver T B A G Q ( C D p e a km a x . a t 2 9 5 n m and T4,T13, the TBA GQ was destabilized by 2–12°C [94]. Aldering et al. [57] reported on the impact of uU and the 3′-amino-uU on the structural dynamics and stability of TBA GQ by substituting them for the (nonstacking) loops T3, T7, and T12. While the uU enhanced the thermal stability of the unmodified GQ (TmC 49°C) at all three T positions, by 2–5°C, the 3′-amino derivative decreased it by 2–7°C. As the amino group replaced the 3′-OH, the internucleotide linkage at this site was 2′-5′ instead of the natural 3′-5′ (this modification of the backbone is named isoDNA; see also later). The authors suggested that the altered backbone caused the destabilization effect with the 3′-amino modifications. On the other hand, none of the loop substitutions changed the topology of the native TBA GQ (CD peak max. at 295 nm and negative peak at 265 nm), and both the unlocked uridine analogs improved the thrombin clotting time [57].

α-Anomeric Sugars. α-Nucleosides differ from the natural β-ones by the inversion of the configuration at the C(1′) anomeric position of the furanose ring (Figure 10, 6) (for a review, see [174]). The T3,G4,T5 that assembles into a parallel-stranded tetramolecular GQ was found to be an inhibitor of HIV infection, especially if the phosphodiester internucleotide linkages were replaced with P-S linkages (see Section 3.3), and its IC50 value was 0.3 µM [175]. When the same oligodeoxynucleotide was built from α-nucleosides, the α(T3,G4,T5), also with P-S linkages, a similar anti-HIV activity was observed, and the IC50 value was 0.5 µM [176]. Since α-oligonucleotides are nuclease-resistant [174], it was claimed that the P-S backbone is mechanistically required for antiviral activity of this oligonucleotide. It was, therefore, suggested that PS-ODN interacts with the highly cationic topositoft he nat iver T B A G Q ( C D p e a km a x . a t 2 9 5 n m and T4,T13, the TBA GQ was destabilized by 2–12°C [94]. Aldering et al. [57] reported on the impact of uU and the 3′-amino-uU on the structural dynamics and stability of TBA GQ by substituting them for the (nonstacking) loops T3, T7, and T12. While the uU enhanced the thermal stability of the unmodified GQ (TmC 49°C) at all three T positions, by 2–5°C, the 3′-amino derivative decreased it by 2–7°C. As the amino group replaced the 3′-OH, the internucleotide linkage at this site was 2′-5′ instead of the natural 3′-5′ (this modification of the backbone is named isoDNA; see also later). The authors suggested that the altered backbone caused the destabilization effect with the 3′-amino modifications. On the other hand, none of the loop substitutions changed the topology of the native TBA GQ (CD peak max. at 295 nm and negative peak at 265 nm), and both the unlocked uridine analogs improved the thrombin clotting time [57].
**Other Modifications of the Sugar Moiety.** Besides LNA, another conformationally rigid sugar analog of dG, the *syn* and *anti*, north- and south-bicyclo[3.1.0]hexane-dGs (Figure 11, 4) were substituted for G14 and G15 of the TBA GQ [189, 190]. Substitution of position 14 did not change the stability and topology of the wild-type TBA GQ; substitution of position 15, however, caused a strong destabilization. Coppola et al. [191] substituted TBA’s thymidines, one at a time, with acyclic thymine analogs, namely, N1-(3-hydroxy-2-hydroxymethyl-2-methylpropyl)-thymine (Figure 11, 5), a sort of unlocked nucleotide. Nearly all TBA analogs were able to fold into similar GQ structure as the native DNA did. Substitution of loops T3, T7, and T12 by the acyclic-Ts slightly stabilized, by 1–4°C, the GQ, relative to the unmodified TBA, whereas that of T4 and T13 slightly destabilized, by 3°C, and that of the T9 significantly reduced the thermal stability of the unmodified TBA, from 50°C to 39°C [191].

**Multiple Modifications of the Sugar Part.** When hemin interacts with certain GQs, the complex will exhibit peroxidase activity. These hemin-binding GQs have recently emerged as important peroxidase mimicking DNAzymes and have been utilized in bioanalytical applications. Li et al. [192] studied the effect of multiple chemical modifications of the sugar part: 2′-O-methyl-, L-nucleotides and LNAs; furthermore, the backbone was also modified with phosphorothioates (P-S). Three GQs with known high peroxidase activity were selected for the experiments, the PS2.M, GTG3TAG3CG3TG3G2, the c-Myc, TGAG3TG4AG3TG4A2, and the EAD2, CTG3(AG3)3A. Results indicated that the 2′-O-methyl modification facilitated the formation of perfectly compacted parallel structures with the highest thermal stabilities among the unmodified and modified GQs and with significantly promoted peroxidase activity of GQ based DNAzymes. The LNA, L-, and the P-S backbone modifications reduced the stability of wild types [192].

GRO29A is a GQ-forming, growth-inhibitory oligonucleotide, whose sequence is T3G3TG3TG3TG3G2GTG2TG3G2TG3G2 [193]. Several sugar and backbone-modified analogs...
of GRO29A have been prepared by incorporating 2'-O-methylribo-, mixed 2'-deoxy, and 2'-O-methylribo nucleosides and backbone-modified P-S analogs. The full P-S, the 2'-O-methyluracil (RNA) 2'-deoxyguanosine-containing chimeric and the full 2'-O-methylribo analogs increased the thermal stability of the GQ of the 3'-aminoalkyl-modified GRO29A, while the other analogs decreased it or no thermal profiles were detected. The P-S and chimeric analogs were found to significantly inhibit proliferation of a number of tumor cell lines, but the 2'-O-methyl RNA analog had no significant effect. Based on molecular modeling, it was concluded that the inactivity of the latter was due to the differences in the groove structure of the GQ, compared to the other's [194].

The effects of multiple substitutions of the TBA DNA by LNA and 2'-O-methyl analogs of G (LNA-G and om2G) on the formation and topology of the TBA GQs have been studied. Results showed that when two or more Gs were substituted with LNA-G or om2G, the 15-mer TBA oligonucleotide remained unstructured in 50 mM K+ The native TBA GQ was unstructured in 50 mM Ca2+, and interestingly, four or more LNA-G or om2G substitutions induced the formation of parallel GQs, which were dimeric forms [195].

Chain Terminal Modifications. "Sweetening" the TBA GQ, that is, attaching various 5'-terminal sugar derivatives to the 15-mer TBA DNA, led to the reduction of the thermal stability of the native GQ by 4.5 to 10.1°C. The sugar moieties were connected to the 5'-DG through a phosphodiester group, for example, the β-D-glucose-C4-OPO2-5'-TBA, β-D-Galactose, -maltose, -lactose, -cellobiose, and β-L-fucose were also appended. Compared to the HO-C4-OPO2- derivatized TBA GQ, not the native TBA GQ, there was a slight increase in the Tm values, 0.7 to 2.3°C, with the added sugar moieties. Attaching carbohydrates to the 5'-end of the telomeric oligonucleotide of TAGGGTTAGGGT could alter the folding topology of the parent GQ. The sugar moieties can well stack with the 5'-G-tetrad [196]. Four cationic amino acids, lysine, ornithine, homo-arginine, and arginine, were covalently attached to the 3'-end of the tetrahymena telomeric repeat DNA sequence TG4T [197]. The cationic residues did not affect the formation of parallel tetramolecular GQ structures, and the 3'-conjugations increased the thermal stability of [TG4T]4 by 3 to 6.4°C in Tm in 0.11 M Na+ [197]. Thermal stabilization by several other terminal modifications of tetramolecular GQs has been reviewed by Doluca and coworkers [4] and more recently by Prokofjeva et al. [15] who analyzed the GQ-forming chain end-modified oligonucleotide aptamers, primarily the Hotoda TGGGAG GQ analogs, in connection with their anti-HIV activity. The anti-HIV activity of GQs is exerted via preventing the viral entry into the target cells by binding to the gpl20 surface glycoprotein of the virus and by targeting the viral integrase or the viral reverse transcriptase within the HIV, as it was reviewed by Musumeci et al. [16]. The 5'-terminal modifications were the dimethoxytrityl, 3,4-dibenzylxoybenzyl, tert-butylidiphenylsilyl, and other aromatic groups. 3'-End modifications also enhanced the aptamer activity, such as the 2-hydroxyethylphosphate, glucose, or the mannose moieties. LNA-G at certain tetrad positions also resulted in potent activity with increased thermal stability of the GQ. No strict correlation was found between the stability and the aptamer activity of the GQs. Terminal modifications, especially the hydrophobic groups, can increase the thermal stability of GQs. The backbone modifications, such as the thiophosphoryl (P-S) internucleotide linkages, increased the resistance to cellular nucleases, which is important for the aptamer activity [15].

3.3. Phosphodiester and Sugar-Phosphate Backbone Analogs

Phosphorothioates (P-S). A sulfur atom replaces one of the nonbonded oxygen atoms of the phosphodiester (P-O) internucleotide linkage (Figure 12, 1) in the phosphorothioate (P-S) linkage (Figure 12, 2). The sulfur can be in Rp or Sp stereo configuration, bringing in the chirality problem. In most cases, however, the isomers are not separated after the synthesis. Sulfur has a larger size than the oxygen and is less hydrophilic. If the P-O linkages are replaced with the random stereo mixture P-S linkages in a DNA or RNA strand, its duplex with an unmodified complementary strand is thermodynamically destabilized relative to the full P-O duplex. If the pure stereo full-Rp or full-Sp isomeric strands are compared, the negative effect of Sp isomers is larger than the full-Rp strands in their duplexes [198]. However, the P-S strands become highly resistant against cellular nucleases, and this property made the P-S oligonucleotides the first generation of therapeutically applicable antisense oligonucleotides, as reviewed by Eckstein [199]. As for the GQs, when all P-O linkages were replaced by P-S in the tetramolecular parallel [TG4T]4, in the 12-mer bimolecular [G4T4G4]2 GQs in Na+ and in the intramolecular antiparallel TBA GQ in K+, the GQs were all destabilized. The P-S substitution of the TBA was studied in more detail by Zaitseva et al. [200] and Prokofjeva et al. [15]. The CD spectrum of the unmodified TBA GQ was not changed by P-S modifications. Presence of a single P-S bond between two G-quartet planes (the G-G bond) led to a significant decrease in GQ thermostability, from 52°C to 45°C in 100 mM K+. On the contrary, modifications in the loop sequences (T-T bonds) either did not alter or stabilized the structure [15].

Interestingly, the P-S substitution had no effect in Na+ and increased the stability in K+ of the intramolecular GQs built from the 18-mer AGG(TTAGG)3 and the AG3(TTAGG)3 [1]. Another sequence, the T7G4T2, forms a parallel-stranded GQ, which has antiviral activity. The tetramolecular structure binds to the human immunodeficiency virus envelope protein gpl20 at the V3 loop and inhibits both the cell-to-cell and virus-to-cell infection. The P-S analog of T7G4T2 was found to be a more potential inhibitor of HIV infection, in vitro [175, 176]. Site-directed spin labeling technique was used for investigating GQ conformations by Zhang et al. [201]. Double nitrooxide labels were attached to the sulfur atom of a P-S linkage at G3 and G15, the G4 and G15, and the G9 and G15 of the A(htel-21) oligonucleotide. The labels only marginally impacted the folding feature of the GQs and could adequately
Figure 12: 2'-Deoxydinucleotides connected by the natural phosphodiester (P-O) (1); the thioosphosphoryl or phosphorothioate (P-S) (2) and the nonionic methylphosphonate (P-Me) linkages (3); the 1,4-disubstituted dihydroxyanthraquinone linker (4); the acyclic (R)-4-aminoobutane-1,3-diol phosphodiester backbone (5); and the pyrene molecule used for loop (6).

distinguish between different topological conformations of the A(htel-21) GQ.

Inversion of Polarity Sites. In natural nucleic acids, the nucleotides are connected via 3'-5' phosphodiester linkages. When this is changed to 3'-3' or 5'-5' linkages, it is called inversion of polarity. In GQ-forming oligonucleotides, such sites can be introduced at three different locations: in non-G-tracts, in G-tracts, and between a non-G-tract and a G-tract. The polarity inversion modifications in GQs have been reviewed by Virgilio et al. [9]. The inversion polarity site-containing GQs, a new class of GQs, have been introduced in 2005, and the first such structures contained 3'-3' or 5'-5' inversion sites in the G-tract [202]. Later, a library of 14 inversed-polarity site-containing TBA GQs have been prepared: seven with 5'-5' and seven with 3'-3' inversions. The former started with the 3'-G5'-5'GTTGGTGGTGG-3' and continued up to 3'-G5'-5'GTTGGTGGG-3'. The latter group started in 5'-G5'-5'GTTGGTGGTGGG-5' and continued up to 5'-G5'-5'GTTGGTGGTGGG-5'. The structures were characterized by NMR spectroscopy. The 5'-5' inversion site in the G-tract prevented the formation of stable GQ structures [203]. Another study showed that, among the T5'-5'T series, the 3'-GGT5'-5'TGGTGGTGGTGG-3' sequence folded into an unusual structure adopting three strands parallel to each other and only one strand oriented in the opposite manner. This modified TBA was more stable than its unmodified counterpart and showed a high thrombin affinity. $T_m$ of the wild-type GQ was 53°C and that of the inversed-polarity analog was 57.9°C [204, 205]. GQ of the 3'-TG5'-5'GGGT (QS55) and the 5'TG3'-3'GGGT (QS33) formed different topologies. NMR analysis revealed that the latter formed a parallel-like fourfold symmetric GQ, while the former possessed twofold symmetry and was characterized as a tetramer antiparallel GQ embedded between two parallel
tracts. The inversed-polarity-containing GQs had reduced thermal stability relative to the native fourfold GQ. \( T_m \) values of 65°C, 47°C, and 53°C were measured for the [TGGGGT]4, the [5'TGG3'-3'GGGGT5']4, and the [3'TGG3'-5'GGGGT3']4, respectively [206]. NMR and CD studies of two polarity-inversed GQs, named Q33 and Q55, formed by the oligodeoxynucleotides 5'-TGGG3'-3'GGTG5'-3' and 3'-TGG5'-5'GGTG-3', respectively, had different arrangement of the glycosidic angles of the residues, which led to a different symmetry and different physical properties of the two GQs. Both GQs assumed right-handed conformation. The Q33 had all G residues in anti glycosidic conformation, while the Q55 had one G-quartet with all-syn residues. Interestingly, the Q55 was more stable by 25°C than the natural counterpart, whereas Q33 was a less stable structure (\( T_m = 52°C \)) [207]. Inversed-polarity sites were also introduced into the TG4 oligonucleotide, which folds into [TG4]4 in K+ solution [208]. The presence of these unnatural sites did not hinder the formation of tetramolecular GQs and in some cases increased the thermal stability of the unmodified one. \( T_m \) of the GQ formed from the native sequence TGGGT was 45°C, that of 5'T3'-3'GGGGT5' was 44°C, and that of 5'TGG3'-3'GGT5' was 49°C; that of 3'T5'-5'GGGT5' was 45°C; and that of 3'TGG5'-5'GGT3' was 72°C [205]. Virgilio's group has also studied 5'TGGGGT5' analogs that contained two or three 3'-3' or 5'-5' inversion sites in the G-run, namely, 5'TGG3'-3'GGG5'-5'GGT3', 5'TGG3'-3'GGG-5'GGT5', 3'TGG3'-3'GGG5'-5'GGT3', 5'TGG3'-3'GGG5'-5'GGT5', and 3'TGG5'-5'GGT3'-5'GGG5'-5'GGT3' [209]. The modified sequences contained either no or only one natural 3'–5' linkage in the G-tract; notwithstanding, all analogs formed stable tetramolecular GQs. Results showed that the 3'-3' or 5'-5' inversion sites affected the glycosidic conformation of dGs and, consequently, also base stacking. This study allowed the authors to depict some generalizations concerning strand arrangements and the glycosidic conformational preference of residues adjacent to inverted polarity sites [209]. The inversion sites could increase the resistance of the TBA GQs against degradation by cellular 3'-exonucleases. Eight TBA analogs have been examined, which contained the inversion site at the 3'-end of the oligonucleotides that also contained an added natural base at the 3'. Some of them contained also a 5'-inversion site also with an added nucleotide. A few TBA analogs not only were much less sensitive towards various exonucleases but also were active apamers and had elevated thermostability. The 5'-G3'T2G3'TGTTG2'T5; G2'-3'3'-A was the most stable with a \( \Delta T_m \) of 12°C, and the A5'-5'-G3'T2G3'TGTTG2'T5; G2'-3'-T was the least stable with a \( \Delta T_m \) of -1°C, relative to the stability of the unmodified GQ. The modifications did not change the antiparallel, chair-type conformation characterizing the wild type [210].

Polarity inversion-site-containing GQs proved more resistant to exonucleases than the wild types [210, 211], and some of them also showed antiproliferative activity [211]. GQ analogs have also been built from (TG4)4, (TG5)4, and (TG4)4 that contained one 3'-3' and two 5'-5' inversion sites. The modified (TG4)4 and (TG5)4 formed very stable monomolecular parallel GQ structures characterized by three side loops containing the inversion of polarity sites. Both contained an all-syn G-tetrad, while the other dGs adopted anti glycosidic conformations. Both GQs showed remarkable antiproliferative activity against cancer cell lines [211]. Oliviero et al. [212] published on polarity site inversion GQs with the aim of obtaining structurally homogenous DNA G-wire nanostructures, using 5'-CGGT-3'-3'GGC-5' sequences. An NMR study performed by Šket et al. on the tetramolecular GQs formed by TG4T and its analogs containing a 5'-5' or 3'-3' inversion of polarity site, namely, the 3'TGG5'-5'G3'T3', 5'T5'-5'G5'T3', and the 5'TGG3'-3'G3'T5', revealed that the parallel GQs had distinct cation-binding preferences [213]. When not only one or more inversion sites are introduced into an oligonucleotide but also the whole sequence is reversed from the 5'-3' direction to 3'-5', Marušič and Plavec [214] introduced the term sequence inversion or sequence. G-rich oligonucleotides of this full sequence inversion could assemble into GQs; however, both the thermal stability and the number of structures formed changed, as it was described for the natural 5'-3' 27-mer G3TAG3CAG3ACACAG3TAG3, a noncoding segment of the human papilloma virus (HPV) type 52, with its two truncated sequences. The reversed sequence of the 27-mer, illustrated as 5'-3', was the G3ATG3CACACAG3ACG3ATG3. For the three reversed sequence-GQs, the \( T_m \) values were lower by 1 to 7°C in K+ and 3 to 7°C in Na+ solution. The inverted sequences showed completely different folding preferences from what the natural sequences did [214].

Another example for the inversed polarities is with hexaethyleneglycol (HEG). A nonnucleotide bridge connects two TG4 oligonucleotides, whose connection was built with inverted polarities: 5'-TG4G3'–p-HEG-p-3'GAG4T5' and the 3'-GAG4T5'–p-HEG-p-5'TG4G3'. In K+ solution, the molecules formed tetramolecular parallel GQs where the two HEG bridges formed two loops. The wild-type tetramolecular parallel GQ of TG4AG had a \( T_m \) of 41.5°C. The HEG GQs had much higher thermal stability: the first analog 76.0°C and the second 75.5°C. The conjugated GQs exhibited elevated resistance in human serum and high or moderate anti-HIV-1 activity with low cytotoxicity. As a result, these conjugated hairpins represent the first active anti-HIV-1 bimolecular GQs based on the TG4AG sequence [215]. A doubly modified GQ containing polarity inversion has also been described: L-residues and inversion of polarity sites have been introduced into five TBA DNAs by Esposito et al. GQs of the all-L-TBA oligonucleotides as well as the all-L minus the first and third TT loops, which remained D-thymidines, folded into left-handed GQs, and, interestingly, thermal stability of the wild type was retained in both modified structures. Two mixed L- and inversed-polarity site sequences also formed left-handed GQs and their stabilities changed depending on the site of modification. The TBA analogs containing L-residues and inversion of polarity sites lost the anticoagulant activity but gained antiproliferative properties against two cancer cell lines [216]. In an NMR study, Esposito et al. used inversed-polarity-(d(TGGGT))s, which also contained an AP (abasic) site, and revealed interesting changes of the anti and syn glycosidic conformation in the G-tetrads of the tetramolecular parallel GQs [217].
**Nonnucleotide Backbone.** 1,4-Dihydroxyanthraquinone and 1,8-dihydroxyanthraquinone linkers (Figure 12, 4) were substituted by Gouda et al. [218] for the loop T nucleotides of the TBA GQ. Single substitutions generally reduced the stability of native TBA ($T_m$ $50^\circ C$ in 100 mM K$^+$). On the other hand, double substitutions by either linker led to stabilizations by 4 to $26^\circ C$. The latter was achieved with the 1,4-disubstituted anthraquinone linker in positions T4 and T13 in the two lateral TT loops. The modifications retained the antiparallel conformations of the native TBA GQ. Majority of the anthraquinone-modified TBA GQs showed decreases in the clotting times, compared to TBA. The aptamers containing the 1,8-linker at G8 or T9 in the TGT loop had improved anticoagulant activity [218]. Gouda et al. [219] also replaced the TT and TGT loops of the TBA GQ with novel 1,5- and 2,6-disubstituted anthraquinones. Single substitutions destabilized, while anthraquinones in two TT loops led to 1–18°C increase in $T_m$ of the parent GQ without changing the original topology. TBA nucleotides have also been substituted with acyclic (R)-4-aminobutane-1,3-diolphosphodiester backbone (Figure 12, 5). Substitution of T7 resulted in a slight increase in $T_m$ with no effect on conformation. Double substitutions of 5G56, or multiple T positions, resulted in significant decreases in the $T_m$ and the $\Delta G_{310}$ values [220]. One, two, and all three TTT loops were replaced by pyrene molecules (Figure 12, 6) of the 21-mer G5(T$_3$G$_3$)$_3$-oligodeoxynucleotide by Rajagopal and Hariharan [221]. In 100 mM KCl, the thermal stability of parent antiparallel/hybrid GQ (positive max. around 290 nm with a strong shoulder/peak near 260 nm, negative max. close to 240 nm in its CD spectrum) of 70.2°C was not changed by a single pyrene backbone, $T_m$ 69.9°C; however, the stability was increased to 74.5°C and 87.2°C for the two and three pyrene backbone-containing GQs, respectively. Concurrently, CD spectra showed the formation of scaffolds with elevated content of parallel strands by the increase of the 260 nm peak. In the deep eutectic solvent of 1:2 choline chloride-urea containing 100 mM KCl, mimicking visous biological conditions, the $T_m$ values decreased from 68.8°C to 60.9, 40.8, and 36.6°C for the one, two, and three pyrene backbone-containing GQs, respectively. CD spectra of the unmodified GQ showed parallel folding, to which the pyrene-modified GQs converted from hybrid form with increasing pyrene content [221].

**4. Synthetic Nucleotides That Only Destabilized a GQ Structure**

**4.1. Destabilizing Base Derivatives**

6-Thioguanine ($\text{s}^6\text{G}$) and 6-Mercaptotpurine ($\text{s}^6\text{Pu}$). The medication 6-thioguanine (Figure 13, 1), used against leukemia and ulcerative colitis, has also been introduced into various GQ-forming sequences, such as the TG$_4$T, TG$_3$T, TBA, and the hel repeats. In K$^+$ the $\text{s}^6\text{G}$ in none of the G positions altered the characteristic CD spectra of [TG$_2$T]$_4$ and [TG$_3$T]$_4$ but sharply destabilized the structures, as Gros et al. reported [222]. $\text{s}^6\text{G}$ was also incorporated into position 13 of the two-tetrad-forming (hel-21)T. The GQ contained a br$\text{s}^6\text{G}$ at position 7, as well, and the effect of $\text{s}^6\text{G}$ on stability was not separately discussed. The overall topology was not changed by $\text{s}^6\text{G}$ [77]. However, $\text{s}^6\text{G}$ inhibited the formation of TBA’s antiparallel, two-tetrad, chair-type GQ when it substituted for G2 in G$_4$$\text{s}^6\text{GTG}$T$_2$G$_2$T$_2$G$_2$ or in G$_4$([TU]$_2$G$_2$[UT]$_2$G$_2$U$_2$T$_2$G$_2$) [223] although the U base alone is not an inhibitor of GQ formation, for example, in G$_4$(T$_3$G$_3$)$_3$ [224], which is only one of the many cases known [5, 8]. The destabilizing effect of $\text{s}^6\text{G}$ was also studied by molecular dynamics calculations [225]. Destabilizing effect of a new purine analog, 6-mercaptopurine, has been described by Radhika et al. [226] based on MD simulations of the GQ structure of (TGGGG)G$_4$ containing a single analog that also caused local distortion of the fold.

6-Selenoguanine ($\text{s}^6\text{G}$). Molecular dynamics simulations of the TBA GQ showed that although one nucleotide of $\text{s}^6\text{G}$ was tolerated in the scaffold, the thermodynamic stability was reduced accompanied by conformational alterations. Two or more such mutations prompted unfolding of the scaffold due to steric clashes in the interior channel of the GQ, which led to the release of the central K$^+$ ion and to the disruption of the structure [227].

7-Deazaguanine ($\text{c}^7\text{G}$). In GQ models, the 7-deazaguanine (Figure 13, 2) was first mentioned in 1992 demonstrating that “regular” GQ structures cannot be formed if $\text{c}^7\text{G}$ replaced a G in the core [228]. Biological functions were also impaired by this analog, as it was shown when it substituted individually for each G-position of the 15-mer TBA DNA [19, 20]. The TBA aptamer binds to and inhibits thrombin. 7-Deazaguanosine in each of the eight tetrad positions significantly reduced the inhibitory activity of TBA. Substitution of G8 of the central TGT loop had a minor effect as demonstrated by the 2-fold increase in the inhibition constant. Later, $\text{c}^7\text{G}$ was also incorporated into the DNA sequences forming tetramolecular architectures, such as the TG$_4$T and TG$_3$T [222]. The modified oligonucleotides formed the same type of parallel structures as the unmodified ones did although the Hoogsteen-type H-bonding could not be formed in the absence of the H-acceptor of N7. Due to the steric requirements of the hydrogen atom of C7, the $\text{c}^7\text{G}$-tetrads became deformed as compared to the full-G tetrads, resulting in changes of stacking. This led to large reduction in the thermal stability of [TG$_2$T]$_4$ by the $\text{c}^7\text{G}$-analogs. In general, intramolecular GQs cannot form or are severely destabilized by a single $\text{c}^7\text{G}$ in the core, which made $\text{c}^7\text{G}$ a “principle-proving” analog [229–232]. The $\text{c}^7\text{G}$ derivative 8-aza-7-deaza-isoguanine (n$^8\text{c}^7\text{G}$) (Figure 13, 3) nucleotide-containing DNAs can form tetrads and also pentads and thus GQ and pentaplex scaffolds. Seela and Kroschel [39] described that the T$_4$(n$^8\text{c}^7\text{G})$_5$T$_5$ molecules self-assembled into GQs in the presence of Na$^+$ or Rb$^+$, whereas they formed pentaplexes with Cs$^+$ cations. Structural stability of the GQ assemblies is not known.
Pyrene-Perylene da. GQ-induced FRET (fluorescence resonance energy transfer) has been studied by substituting loop adenosines at position 8 with a tethered FRET pair in A(3-tel-21), A(G)ₙTTPyA(G)ₙTTPerA(G)ₙTTA(G)ₙ, where the n changed from 2 to 4, Py is pyrene (donor), and Per is perylene (acceptor) group (Figure 13, 4, 5). The substituted bases did not hinder the formation of intramolecular 2-, 3-, and 4-tetrad GQs in K⁺; however, they marginally destabilized them by 1-2°C, relative to the unmodified control GQs [233].

6-Methyl-isoxanthopterin. A fluorescent G analog 6-methyl-isoxanthopterin (6MI, Figure 13, 6) has been incorporated into the central TGT loop (position G10) and also at the middle G-tetrad (position G13) of the GQ-forming G₃T₃G₃TGGGTG₃G₃ [234], whose DNA is part of the promoter region of c-MYC oncogene. The loop modification slightly decreased the thermal stability of wild-type GQ, while the tetrad modification significantly did the same. CD spectrum of the wild-type and the modified ones displayed a strong positive band at 263 nm and a minor peak at 298 nm, which the authors, probably mistakenly, called antiparallel basket and/or chair topologies. The fluorescence intensity of the loop-modified GQ was greater than the tetrad substituted, as the authors expected from the extent of stacking interaction, which is larger in the G(6MI)G sequence than in the T(6MI)T sequence. The fluorescence emission intensity decreased when 6MI was incorporated into single-stranded oligonucleotide and decreased further in the double-stranded oligonucleotide, demonstrating that its fluorescence emission intensity can be used to probe the microenvironment of 6MI. The 6-methyl-isoxanthopterin was also substituted for the T₄ (loop) of (G₃T₃)₂G₃ oligonucleotide to develop fluorescent detection methods. The wild-type sequence folded into parallel GQ in K⁺. The substitution did not change the original topology but destabilized the GQs. Tₘ of the (G₃T₃)₂G₃ GQ, measured in 10 mM KCl, was 90°C, which was reduced by 12°C by the base analog [124].

Figure 13: Structures of 6-thioguanine (1); 7-deazaguanine (2); 8-aza-7-deaza-isoguanine (3); pyrene (4) and perylene (5) tethered to C8 of guanine; 6-methyl-isoxanthopterin (6); zebularine or 2-pyrimidinone-riboside (7); and 4-thiouracil (8).
2-Pyrimidinone and 4-Thiouracil Nucleosides. Mammalian telomeric DNA is transcribed into RNA that contains the 6-mer uuaggg repeats [235]. Sequence analogs of this, like the uagggu, and those containing analogs of uridines, the zebularine (ribonucleoside of 2-pyrimidinone) that does not have the O4 atom, and the 4-thiouridine (s^4U) (Figure 13, 7, 8), in which the sulfur atom cannot develop H-bond, have been prepared. The tetramolecular parallel GQs formed by the wild types have been radically destabilized by s^4U: T_m value of 79°C for [uagggu]_4 was reduced in [uagggss^4U]_4 to 49.4°C [236]. Mendelboum and coworkers [237] prepared 4-thiouracil-2'-deoxyuridine- (s^4dU-) containing TBA oligonucleotides, in which the base analog substituted for thymines of the loop sequences. Replacement of four thymines by s^4dU in the 15-mer resulted in G_s^4UTG_s^4UGs^4UG_s^4UG. The TBA analog showed increased anticoagulant and antithrombotic properties relative to the unmodified TBA. The increased activity was explained by the altered properties of s^4dU as compared to thymine. The substitutions rendered the molecule more hydrophobic, which might have been preferable for the aptamer-thrombin interaction; furthermore, the s^4dU-containing oligonucleotides have been described as highly resistant to cellular nucleases. This aptamer analog contained more than 26% of thiolated nucleotides, and thus it must have been more stable in a biological environment than its unmodified counterparts. Thermodynamic stability of the s^4U-TBA GQ was not specified [237].

4.2. Destabilizing Substitutions of the Sugar and Backbone Moieties

Acyclic Threoninol. An acyclic sugar analog, the acyclic threoninol (aTNA)-guanine (Figure 14, 1), has been substituted by Zhou’s group [238] for each guanine, one by one in TG_4T. Based on the CD spectra, all the modified oligonucleotides could form GQ structures and only the G3-substituted TGG_GT’s conformation differed from the tetramolecular parallel GQ structures formed by the natural counterpart, [TG_4T]_4. This modified oligonucleotide built multiple scaffolds. Thermal stability of [TG_4T]_4 was reduced by the aTNA-G in each case, and the modification at the 5’- and 3’-terminal G-tetrads was the most detrimental to the stability [238].

Dibenzyl Linker. TBA is primarily known as an anticoagulant aptamer. Another less known biological activity of TBA is its anticancer potential, which is rather hindered by the anticoagulant action, as reported by Scuotto et al. [239]. They found that replacing one residue of the TT or TGT loops with a dibenzyl linker (Figure 14, 2), with which seven new GQ-forming TBA sequences were created, could maintain the antiproliferative activity over the anticoagulant activity. Most T-substitutions only slightly affected the thermal stability of the wild-type TBA, except the T9-modification that reduced T_m of TBA (50.7°C in 90 mM KCl) by 15°C. The T13-modified analog possessed selective antiproliferative activity, while the T12 analog retained the potent anticoagulant activity of
the unmodified TBA. Structural analyses indicated that the different localization of the two benzene rings of the linker was responsible for the loss of the antithrombin activity of the T13 analog.

Methylphosphonate. In the methylphosphonate (P-Me, Figure 12, 3) internucleotide linkages, a methyl group replaces one of the nonbonded oxygen atoms, and thus the internucleotide linkage becomes uncharged, neutral. This causes the ion-solvating water spine-perturbed along the backbone, resulting in the destabilization of a folded structure. With P-Me linkages-containing [TG₄T₁₄], [G₄T₄G₁₄], and the TBA GQs, no thermal transition was detected, and the AG₃(TTAG₃)₃ GQ was powerfully destabilized [1].

isoDNA. The name refers to the modification of the internucleotide phosphodiester linkage, in which the natural phosphodiester bond, 3'-O-OP(O')-5'O, of nucleic acids is changed, for instance, to 2'-O-OP(O')-5'O, which is the 2'-5' isoDNA (Figure 14, 3). The fully 2'-5' linked isoTBA formed unimolecular antiparallel GQs in the presence of K⁺ ions, similar to how the unmodified, 3'-5' linked TBA oligonucleotides did. The isomeric TBA had lower thermal stability than the native one. The TBA₃⁻T₄₅ value in K⁺ was 52°C, while the isoTBA GQs was 37.1°C. When the T7 and T9 were replaced by U nucleotides, the stability increased up to 45°C. On the other hand, the isoTBA exhibited higher stability against exonucleases and were capable of retaining the biological function of the native TBA, that is, slowing down the process of blood clotting [240]. The antiparallel chair-type TBA GQ, whose nucleotides are connected via 3'-5' phosphodiester linkages, folded into parallel GQ when the 2-3-2-nucleotide-long loops (TT...TGT...TT) were shortened. The 2'-5'-isoTBA analog, interestingly, retained the antiparallel topology with the shorter loops such as 2-2-2 or 1-3-1 or even 1-1-1 nucleotides of the resulting G14- or 11-mer GQs. Thermal stabilities of the GQs were, however, reduced by the introduction of shorter loops into the 2'-5'-isoTBA sequences, from Tₘ of 48°C for the unmodified 3'-5' TBA to 34°C for the isoTBA(232) and down to 20°C for the isoTBA(111) in 100 mM KCl [241].

5. Natural Base Lesions and Epigenetic Modifications in GQ DNAs

Exogenous and endogenous chemicals and radiation cause many types of lesions in cellular DNA. The nucleotides are modified by different types of reactions, such as oxidation, alklylation, and hydrolysis. These alterations are widespread and play an important role in changing the physiological states of cells and can thus lead to various diseases. Majority of genetic impairments are believed to originate from oxidative processes, which are the basis of mutation, aging, cell death, and carcinogenesis [242–244]. Reactive oxygen and nitrogen species (ROS and RNS) arising endogenously and originating mainly from the cell's aerobic metabolism also contribute to the age-dependent diseases [244–246]. There are ~80 known DNA defects that can form upon the initial attack of ROS and RNS [247]. Among the exogenous effects, the damage can arise from radiation, directly from the ionizing energy or indirectly from hydroxyl radicals that form by the ionization of the solvation shell around the DNA [248, 249]. The radical cation (electron ionization hole) can travel hundreds of Angstroms by hopping before being entrapped, preferentially by purine bases [250, 251]. Oxidation of purines initially leads to 8-oxo-7,8-dihydroguanosine (oG), 8-oxo-7,8-dihydroadenosine (oA), and 8-oxo-7,8-dihydroinosine (oI) in RNA, in DNA, and in the nucleotide pool [244]. The effects of natural base lesions on the structure and stability of double-stranded DNA models have been extensively studied (see [252–258] and the references therein). Majority of the lesions are promutagenic and procarcinogenic if not repaired in due time. The base excision repair (BER) pathway, regulated by many different types of enzymes including DNA glycosylases, abasic endonucleases, phosphodiesterases, DNA polymerases, and DNA ligases, is responsible for the accurate removal of the lesion and the restoring of the original state of the double-stranded nucleic acids [243]. Due to imperfections of the repair system, the lesions that are not repaired can severely damage the DNA structure where they are formed. Base damage also occurs in noncanonical DNA structures such as the GQs. Studying the lesion-damaged GQ structures using various GQ models is less than a decade old. A few cases are known only about repairing damaged GQs, such as the N6-methylguanine (m6G) [259] and the further oxidized derivatives of 8-oxoguanine (oG), but oG is not among them [260].

Using GQ models, the oG, oA, and oI (8-aminoadenine), m6G, G- and A-abasic (AP) sites, hm5U, hypoxanthine (I), xanthine (X), and the cyclobutane thymine dimers have been investigated for the effect on GQ stability. Most natural DNA damage is destabilizing both in canonical and in noncanonical nucleic acids; a few, however, do not affect or even stabilize the GQ fold. The stabilizing lesions are the oA and oI, only if located in the loops of the telo-GQs, and also oG if located in the tetramolecular DNA GQs of [TG₄T₁₄] and [TG₃T₄₅]. hm5U, with its minor effects, also belongs to this group of analogs as well as the AP site when replacing a single-base loop in GQs of the T(G₃T₄)₄ and (G₂AP₂)₂.

5.1. Natural Base Lesions That Stabilize or Destabilize, Depending on the GQ Structure

8-Oxo-2'-deoxyguanyline (o8G). Guanine has the lowest redox potential among the four DNA bases [251]; therefore, guanine is the major site of oxidation in DNA by ROS, the reactive oxygen species [261]. Among the numerous oxidative derivatives, the 7,8-dihydro-8-oxoguanine (8-oxoguanine, oG, Figure 15, 1) was found to be the major primary product in vivo [262–264]. Since in vivo levels of oG were found to be safely measurable, being between 0.3 and 4 oG per 10⁶ G, oG has been used as a marker of oxidative stress of cells [265, 266]. oG is even more prone to oxidation than G as its redox potential is even lower [267]. Further oxidation of oG results in guanidinohydantoin and spiroiminodihydantoin, among other minor products [268, 269]. Consecutive runs
of guanines, as those in the potential GQ-forming G-rich sequences, further lower the redox potential of guanine since these runs act as sinks for the oxidative damage [270]. The glycosylases known to repair o\(^8\)G in duplex DNA, such as OGG1, NEIL1, and NEIL3, cannot remove o\(^8\)G from DNA GQs, although these enzymes do repair guanidinohydantoin and spiroiminodihydantoin in GQs [260]. Therefore, the unrepaired, persistent o\(^8\)Gs can ruin the GQ structure. This has been suggested to lead to telomere shortening and finally to cellular senescence [271]. The effect of o\(^8\)G was, however, recently proved by Fouquerel and coworkers to be not unambiguous, as the destabilized, partially unfolded GQ could promote telomerase activity that could lead to telomere extension [272, 273]. Formation of o\(^8\)G in GQs motivated wide-ranging investigations.

The 8-oxoguanine was first studied with tetramolecular parallel GQ structures by Gros et al. [222] using the [TG\(_4\)T\(_4\)]\(_1\) and [TG\(_5\)T\(_4\)]\(_1\) models. These GQs assemble from four TG\(_4\)T and TG\(_5\)T strands, respectively, and therefore the presence of a single modified base in one strand forms a modified tetrad in the GQ. Incorporation of a single o\(^8\)dG into each G-position of TG\(_4\)T and TG\(_5\)T did not hinder the formation of tetramolecular parallel GQs. Interestingly, o\(^8\)dG proved to be a stabilizing modification, especially when incorporated at the 5\(^{\prime}\) position of the G4 or G5 sequence, as characterized by the \(T_{1/2}\) values. As the preferred glycosidic conformation of o\(^8\)dG is syn (Figure 15, 1), similar to the majority of 8-modifications of G [63], and in most cases in the parallel GQs all nucleosides have anti glycosidic torsion angle, o\(^8\)G was supposed to destabilize the structure. Furthermore, in the 6,8-diketo tautomeric form of o\(^8\)G [274], the donor-acceptor arrangement of the Hoogsteen H-bonding changes, which again would have predicted the destabilization of the parallel GQ by the o\(^8\)dG tetrad. The observed stabilization effect was finally explained by strong stacking interactions with the neighboring anti G-quartet. As it turned out later, the elevated stability was observed merely with these two tetramolecular scaffolds. There was another experiment for tetrad formation from o\(^8\)G nucleotides when o\(^8\)dG replaced G in the four 5\(^{\prime}\) positions of the GGG triplets (in bold, underlined) in GGGTG\(_{2}\)G\(_{2}\)GG and results were the opposite, destabilization. This oligodeoxynucleotide folded into an intramolecular parallel GQ, possibly due to the single-base loops. Thermostability of wild type was substantially reduced

![Figure 15: Structures of the syn 8-oxo-2'-deoxyguanosine (1); syn 8-oxo-2'-deoxyadenosine (2); anti 5-hydroxymethyl-2'-deoxyuridine (3); the tetrahydrofuran abasic site (4); anti O6-methyl-2'-deoxyguanosine (5); hypoxanthine (6); and xanthine (7).](image-url)
by $\delta^8 dG$, by $10^\circ$C in Na$^+$, and no unambiguous $T_m$ value could be determined in K$^+$ where an even larger destabilization was apparent. The circular double H-bonded scheme was assumed to consist of NIH-O8G and O6G-HN7 [275].

Reduced stability and conformational changes were observed when a single $\delta^8 dG$ was incorporated into the GQ-forming htel oligonucleotides. Szalai et al. [276] used the 25-mer model A(htel-21)TGT and found that $\delta^8 G$ at 5’ positions of the GGG triplets retained the intramolecular antiparallel topology of the wild type, whereas $\delta^8 G$ in the middle positions of the triplet caused formation of multiple folds. With the 24-mer hybrid GQ of TT(GGGTTA)GGGA, Lech et al. [80] detected significant destabilization of the wild type even in cases when the syn dGs were changed to $\delta^8 dGs$ in positions 3, 9, and 15 (all 5’ positions in the GGG triplets) and the $\Delta T_m$ values ranged between −9.8 and −18.5°C. Concurrently, a mixture of topologies was formed. $\delta^8 dG$ was also incorporated into all 12 dG positions (syn and anti) one by one, into the GQ-forming htel-21, Gt5(TAGTGA)$_3$, by Sagi and coworkers [26]. The study revealed that the GQ structures were substantially destabilized by enthalpy-driven effects both in Na$^+$ and in K$^+$ solutions. The negative effect was position-dependent and varied with the cation used. When the single $\delta^8 dG$ was located in the two terminal tetrad, the Na$^+$-stabilized basket-type antiparallel topology was also retained in K$^+$, contrary to the wild-type htel-21, which forms a mixture of folds, hybrids, and K$^+$-antiparallels in K$^+$ [68, 101, 277, 278]. The middle tetrad substitutions by $\delta^8 dG$ caused the largest reductions in stability (−19 to −26°C in Na$^+$ and −27 to −30°C in K$^+$), as observed before with other base mutations [279, 280], and various folds were observed too. The damaging effect of $\delta^8 dG$ on the stability of monomolecular htel GQs was surprising in light of its stabilizing effects with tetrameroloc GQs [222] (the $\delta^8 G$ triggered only marginal destabilizations in deoxyoligonucleotide duplexes and the effect was also length-dependent: with duplexes of 15 nucleotides long, even a small stabilization was observed; see references in [261]). In an $\delta^8 G$-GG-G-G tetrad, the 8-substitution changes the Hoogsteen-type circular H-bonding pattern. In the 6,8-diketo tautomeric form of $\delta^8 G$ [274], the N7 is a hydrogen donor, instead of acceptor, and therefore only a single H-bonding pattern remains when $\delta^8 G$ is present in the tetrad. The weakened H-bonding and the stacking changes induced by the altered tetrad can result in the extensive destabilization of the GQ.

As $\delta^8 dG$ contained in htel GQ was found not to be a substrate of the glycosylase enzymes that repair $\delta^8 dG$ in duplex DNA, such as the hOOG1 and NEIL enzymes [260], $\delta^8 G$ can become a persistent lesion in GQs. Surprisingly, this persistence was without any immediate negative consequences on the stability of telomere complexes in vivo (see references in [281]). An and coworkers have built a 31-mer telomeric GQ model from TAGGG(TTAGGG)$_3$,TT, which contained five G-triplets, instead of the usual four, in which the dGs at the 5’-positions of 3, 15, and 27 were replaced by $\delta^8 dG$. These modified oligodeoxynucleotides folded into hybrid-1 or hybrid-2 GQ form. They observed that the folded structures could effectively accommodate a single $\delta^8 dG$ by looping out the damaged G-tract and allowing the other four G-triplets to adopt the hybrid fold. This change caused only a minimal negative impact on the stability of the GQ [281], contrary to the results obtained with the four G-tract telomeric GQ models in vitro. Considering the availability of additional G-tracts in the telomeric complexes, the authors explained how the persistent presence of a lesion can exist without immediate destabilization effect on the DNA-protein complex in vivo.

Another way was also found to offset the intense destabilization caused by a single $\delta^8 dG$ of the GQs built from the four TTAGGG repeat htel sequences and this was by the proper site-specific incorporation of another modified base, specifically xanthine (2,6-dioxopurine, X, Figure 15, 7). Xanthine is another natural derivative of guanine (see in the section Xanthine). X can also be assumed to be a destabilizing analog due to the loss of H-bonds if incorporated into a G-tetrad, similarly, as it was found with hypoxanthine (I for inosine) (see under section Hypoxanthine). The outcome of the $\delta^8 dG$-X double modification depends on the positions of the modified nucleotides. Benz and Hartig [282] incorporated first two $\delta^8 G$ and two X nucleotides into the TA(htel-21)T sequences, which folded into Xo$^8 G$:Xo$^8 G$ tetrads-containing GQs. The modified tetrads destabilized the wild-type GQ and the topology depended on the arrangement. The GQ either remained in the basket-type antiparallel or changed into parallel form:

(i) wild-type: TA GGG TTAGGG TTAGGG TTAGGG T
antiparallel
(ii) $\delta^8 G$(3):X(11):X(15):$\delta^8 G$(23): TA OGG TTAGGXX TTA$\delta^8 G$ GTTA$\delta^8 G$ T antiparallel
(iii) $\delta^8 G$(3):X(9):$\delta^8 G$(15):X(21): TA OGG TTTX$\delta^8 G$ TTA$\delta^8 G$ TTAXG T parallel

Later, Cheong et al. [283] found the positions for $\delta^8 G$ and X that could compensate for the negative effect of $\delta^8 G$ on the stability of the wild type and which also retained the original topology. They inserted the G:G:X$^8 G$ tetrads into a series of TT(htel-21)A deoxyoligonucleotides. Conformation of the modified GQ scaffolds remained similar to the original hybrid-1 fold. Three of the modified GQs were destabilized, by 3°C to 12°C in their $T_m$; two had similar stability as the wild type had (56°C), and the G(3)$\delta^8 G$(9):X(17):G(21) tetrad proved to be stabilizing, by 2°C in its $T_m$ relative to the wild type. Cheong and coworkers [284] could also reverse the polarity of H-bonds in a tetrad while the original folding topology of the wild type was preserved.

The effect of $\delta^8 dG$ has been well demonstrated to depend on the secondary structure of nucleic acids [26, 285, 286]. Recent studies provided additional results, such as the comparison of the effect of $\delta^8 dG$ in triplexes and GQs [100], the involvement of $\delta^8 dG$ in the stabilization of “guanine-vacancy-bearing” GQs [287], or a comprehensive study on the complex effect of metal ions and cosolutes on the topology of the A(htel-21) GQ when both $\delta^8 G$ and xanthine were contained in the tetrads [288]. 8-Oxoguanine in GQ loops accommodated well and did not have much effect on
the stability [281]. The Zn(II)-porphyrin complex-induced oxidation of guanines in the GQ of TA(htel-21) yielded $^5G$, and its further oxidized product, the spiroiminodihydantoin. The oxidized bases prompted structural rearrangements of the parallel and hybrid TA(htel-21) GQs into an antiparallel-like conformation [289], such as what was observed earlier with $^8O$G-containing htel-21 GQs [26].

8-Oxo-2'-deoxyadenosine ($^8dA$). The first study with 7,8-dihydro-8-oxo-2'-deoxyadenosine (8-oxoDA or $^8dA$, Figure 15, 2), another oxidative natural base lesion, was carried out by Esposito et al. [290] and Petraccone et al. [291] using the tetramolecular parallel GQ DNA models of [AGGGT]$_4$ and [TAGGGT]$_4$. All the modified oligonucleotides formed the same parallel-type tetramolecular GQs as the unmodified two oligonucleotides did. The $^8dA$ substituting for $dA$ nucleotides in both structures decreased the thermostability of the parent GQs by 14°C and 8°C, respectively, according to absorption-based $T_m$ measurement, and the drastic negative effects were supposed to originate from the formation of the $^8dA$-tetraads [290]. Calorimetry, however, provided different results: the [AGGGT]$_4$ GQ was not destabilized by the analog, formulated as 8-hydroxy-$dA$, $^{0h}dA$ [291].

Among the htel GQ models, $^8dA$ was first incorporated into the A(htel-21) GQ by Aggrawal et al. [292]. The analog moderately stabilized, did not affect, or slightly destabilized the GQ depending on the position of substitution. Single, double, and triple substitutions by $^8dA$ were also investigated. In Na$^+$ the average $\Delta T_m$ of four single substitutions was $\sim 0.5$°C above the wild-type's $T_m$ of 60.1°C, $\Gamma$°C in $\Delta T_m$ of three double modifications, and 4.1°C of the triple. In 110 mM K$^+$ solution the respective values were 2.5°C, 2.7°C, and 8.9°C. The presence of $^8A$ in the loops of A(htel-21) GQ did not change the intramolecular antiparallel conformation in Na$^+$, but in K$^+$ multiple folds were shown by the CD spectra. The stabilization by $^8A$ was explained by the tight binding of K$^+$ into the pocket formed by the O8 of $^8A$ and its loop [292]. In another study [293], the effect of $^8A$ was compared with the effect of other natural loop lesions, such as the adenine abasic (A/AP) site and 5-hydroxymethyluracil (hm$^5$U) in the GQ of A(htel-21). $^8A$ stabilized the Na$^+$-basket and the K$^+$-stabilized folds by 1.7°C and 1.1°C, respectively, as averages of the effects at the four sites. The hm$^5$U (Figure 15, 3) hardly affected the stability, while the AP site (dSpacer; Figure 15, 4) destabilized the GQ structure [293].

5-Hydroxymethyluracil (hm$^5$U). One of the main natural oxidation products by ROS of the DNA thymine is hm$^5$U (Figure 15, 3). The ten eleven translocation (Tet) enzymes oxidize the epigenetically important base 5-methylcytosine stepwise to 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. Pfaffenereder et al. [294] found that Tet enzymes also oxidized T to hm$^5$U in mouse embryonic stem cells. The hm$^5$U also forms by ionizing radiation and oxidation of the deaminated m$^5$C [295]. hm$^5$U was first incorporated into the $^5'T$ of TGGGT oligonucleotide that readily formed the tetramolecular parallel GQ. hm$^5$U was modeled as forming an hm$^5$U-tetrad via an extra H-bond through the 5-OH group [296]. Virgilio et al. [297] replaced each thymine in the three loops of the TBA DNA sequences by hm$^5$U. All six sequence analogs retained the ability of the native TBA oligonucleotide to fold into the antiparallel, chair-like, two-tetrad GQ. hm$^5$U did not affect or slightly, by 1–3°C, increased the thermal stability of the native GQ, except for the hm$^5$U in position 9 of the central TGT loop, which increased $T_m$ of the native by 6°C. All TBA analogs showed decreased affinities to thrombin [297]. Effect of hm$^5$U on the stability and conformation of htel GQs was described in 2015 by two laboratories [293, 298]. Both groups found that the replacement of a T of a TTA loop by hm$^5$U resulted only in negligible effects on stability and did not affect the intramolecular topology of the native GQ. Sagi and collaborators [293] replaced each of the six T nucleotides in A(htel-21), the AGGG(TTAGGG)$_3$ and the $\Delta T_m$ values, determined by absorption-based thermal melting profiles ranging from −1.2°C to 0.5°C in Na$^+$ and from −0.5°C to 0.8°C in K$^+$ solutions, both containing 0.169 mM of the cations. Virgilio et al. [298] observed only stabilization by hm$^5$U when six thymines were substituted either in A(htel-21) or in the 26-mer (TTAGGG)$_2$ TT DNA GQs. With the former GQ, the $\Delta T_m$ values ranged between 0.5 and 2.4°C. $T_m$ values were determined here by calorimetry in 100 mM K$^+$ solutions.

Abasic Site (AP Site). Purine and pyrimidine abasic sites (AP sites) are among the most frequent lesions in cellular nucleic acids formed via spontaneous base loss, mainly depurination, and as intermediates in the enzymatic base excision repair process of various other base lesions [299]. Thousands of purine bases are released from DNA in every human cell daily and the resulting AP sites are highly mutagenic if they are not repaired. Natural AP sites occur in equilibrium states between the hemiacetal and the aldehyde forms. Due to the aldehydes, the AP sites are unstable leading to DNA chain breaks through beta-elimination [299]. Therefore, mostly stabilized forms of AP sites, primarily the tetrahydrofuranyl analog (dSpacer, Figure 15, 4), have been used in studies of abasics with canonical, double-stranded oligodeoxyribonucleotide models (e.g., [252, 253]). With noncanonical models, the effect of abasic lesions was a rather unexplored area until up to a decade ago [20, 300, 301]. With GQs, the well-known tetramolecular model assembling from TG$_3$T was first studied in 2010 by Esposito et al. by replacing the dGs by dSpacer AP sites (G/AP) [302]. In the same year, a natural model, the GQ of a human telomeric repeat sequence $G_1(TTAGGG)_3$, the htel-21 was investigated for the effect of loss of G in G-tetrads by Sagi and coworkers [303]. Later, the GQ of the A(htel-21) was used as a model by Fujimoto’s group [304]. This study also examined the effect of the crowding environment on abasic scaffolds. In 2012, the TA(htel-21) GQ was the abasic model of Virgilio et al. [305]. The main and shared conclusions of these studies were that the presence of a single G/AP in the G-tetrads did not hinder the formation of the tetramolecular or the monomolecular scaffolds. On the other hand, a single G/AP significantly decreased the thermal and thermodynamic stability of the parent GQ in a position-dependent way. With the three-tetrad htel scaffolds, the
substitutions in the middle tetrad led to the largest reduction of stability. For instance, in K+ a single G/AP incorporated into the two outer G-tetrads of the three-tetrad GQ of htel-21 decreased $T_m$, of 73°C (Δ$T_m$ = −6.44 kcal/mol) of the wild type by 10–17°C and by 1.76–3.43 kcal/mol in free energy changes (ΔΔ$T_m$) and by 21–26°C in the middle tetrad. (Due to the formations of multiple folds in the middle tetrad-substituted sequences, no thermodynamic data were calculated from the absorption-temperature profiles.) In Na+ the wild-type GQ ($T_m$, 68°C, Δ$T_m$ = −4.75 kcal/mol) was destabilized by 3–17°C and 0.56–2.92 kcal/mol in the outer tetrads and by 12–23°C and 1.99–3.8 kcal/mol in the middle tetrad [303].

The substitution of the loop-T4 sequence of G$_4$T$_4$G$_4$ by AP sites [306] did not prevent the oligonucleotides from assembling into a dimeric, fold-back GQ either in Na+ or in K+, although thermal stability of the unmodified GQ was found to be reduced. With the TBA GQ built by the GGTTGGTTGGTTGG, in which the bold-underlined Ts were substituted with an AP site, one by one, the AP site in T9 considerably destabilized the native TBA GQ, whereas, the AP sites in the other positions moderately stabilized the unmodified TBA [307]. Depurination of a TTA loop (A/AP) of A(htel-21) also reduced the structural stability of the native GQ. The Δ$T_m$ values ranged from −4.0 to −6.7°C in Na+ and from −0.2 to −6.5°C in K+ [293, 308]. These were much less drastic effects than those caused by the G/AP sites in the tetrads of the same GQ model [303]. Folding topology was, however, changed by the loop A/AP sites in a position-dependent way [293, 308]. The wild-type A(htel-21) GQ is polymorphic in K+ consisting of hybrid-1 and hybrid-2 as well as K+ -stabilized antiparallel structures [309]. The loss of a guanine in any G-tetrad shifted the equilibrium towards the antiparallel type [303], whereas, with two loop A/AP abasic GQs (A/AP in the first loop, abbreviated as ap7, and in third loop, ap19), a shift towards the formation of higher population of parallel strands in the mixture was observed. This means the shifting of conformational equilibrium towards higher concentrations of the hybrid folds. Actually, ap19 turned out to be a pure hybrid-2 type fold and the ap7 a hybrid-1 architecture, although the latter formed clearly only at high, millimolar strand concentrations. Folding topology of the wild-type htel-21 and A(htel-21) GQs has been found to depend on the oligonucleotide strand concentrations and to change qualitatively above 2 mM of strands [310–313]. Polymorphism increased when the A/AP was located in the middle loop at position 13, and this was the ap13 structure(s). The A/AP site in the GQs of ap7 and ap19 was located in propeller-like loops. The abasic GQs transformed easier than the wild type into parallel GQs under dehydrating conditions (57% ethanol). The GQ of A(htel-21), in which a loop adenine was replaced by an AP site, the ap7, ap13, and ap19, folded into parallel GQ in K+ even in the absence of ethanol. This study [308] showed that the loop adenine AP sites are potent tools to fine-tune the folding arrangements of the telomeric GQs. The loop-abasic site-induced "monomorphism" has been confirmed by Wu et al. [314], who used this result for a sensor platform. They stated that an AP site replacing a loop adenine can extremely narrow the structure distribution to a specifically monomorphic GQ conformer, depending on the position of the AP site.

Loop thymines have also been changed to AP sites. Rachwal and coworkers. [129] replaced the loop thymines in the propeller-type single-base loops of the monomolecular parallel GQ assembled from the 17-mer T(G$_3$T)$_4$. The substitution did not change the conformation; however, surprisingly, it increased the thermostability of the structure [129]. Another study by Esposito et al. [315] found that single AP sites replacing loop Ts in the GQs of the 16-mer aptamer (GGGT)$_5$ (T30923 or T30695) were not able to affect significantly the conformation and stability of the original structure. The original folding topology was a 5’-5’ dimer of two stacked parallel GQs, each containing three G-tetrads and three single-thymidine reversed-chain loops. This aptamer GQ has been reported to exhibit anti-HIV activity by targeting the HIV integrase. With four such GQ analogs, authors could shed light on the steric interaction of the GQ with the integrase [315]. (A former study [20] also showed that the AP site-containing TBA GQs have altered biological activity: an AP site substituting for any G residue of a G-tetrad in the TBA GQ significantly reduced the inhibition of throbmin. Substitution for either T4 or T13, which stack on the tetrad, also significantly diminished the inhibitory activity of the TBA GQ [20].) Sekridova et al. [316] studied the single-nucleotide loop-containing, two-tetrad GQ of G$_2$AG$_2$G$_2$AG$_2$ from the human ALU-repeat fragment, which was characterized by parallel topology and high thermodynamic stability. The loop bases were replaced with the flexible, nonnucleotide triethylene-glycol (teg) or a tetrahydrofuranyl AP site. The triethylene linker in (G$_2$tegl)$_2$G$_2$ decreased $T_m$ of the native GQ significantly, from 74°C (in 100 mM KCl) to 50°C, while the AP site in (G$_2$AP)$_2$G$_2$, surprisingly, increased $T_m$ considerably, up to 85°C. Furthermore, AFM (atomic force microscopy) data suggested that the two-tetrad GQ with the abasic loops was prone to dimerization in 10 mM KCl. The dimers were supposed to be stacks of monomolecular GQs. At high concentration, 200 mM, of KCl, this loop-abasic GQ formed higher-order aggregates, that is, short nanowires of ~20 nm length, that were long G4-stacks [316]. The loop-abasic sites have been used in several other GQ model studies, such as in the ribo analog of the telomeric DNA and TERRA [169] and also with A(htel-21) [313]. Heddi et al. [317] investigated the formation of GQs containing 4n-1 guanines in the core and synthesized, among others, the T$_5$GGGT(G$_3$T)$_3$ oligodeoxynucleotide, where the G is for the G7/AP site in the middle tetrad of the three-tetrad GQ. The G/AP site drastically reduced the thermal stability of the unmodified parallel GQ, down to 51.2°C from >85°C in 60 mM K+ [318].

5.2. Destabilizing Natural Base Lesions

O6-Methylguanine (m$^6$G). O6-Methylguanine is a major base lesion of nucleic acids in vivo [319], which is promutagenic and procarcinogenic [320]. The anti m$^6$G (Figure 15, 5) destabilized the tetramolecular GQ [TG$_4$T$_4$] in a position-dependent manner in Na+ by forming m$^6$G-tetrads, which did not change the parallel characteristics of the CD spectrum.
of the wild-type GQ [222]. Incorporated in place of each G at the 5'-end G-triplet of AG₄(TTAG₄)₃, which represented each of the three G-tetrads, the mG reduced the thermal stability of the unmodified GQ. The largest destabilizing effect was observed with the middle tetrad substitution, both in Na⁺ and in K⁺. CD spectrum of the unmodified GQ in Na⁺ was only slightly modified by mG. In K⁺, the unmodified oligonucleotide formed hybrid-type mixed folds but the spectra of the outer tetrads-modified GQs looked like their Na⁺-spectra. Spectrum of the middle tetrad-modified structure was qualitatively different from the others. The decreased thermal stability might have been the consequence of missing cation coordination, the disrupted circular H-bonding, and the reduced stacking interactions [321]. Bulky substituents linked to O6 of G, such as the O6-p-nitrobenzyl group, completely hindered the formation of GQ from the TBA oligodeoxynucleotide when substituted at G1. Adding Na₃S₂O₄ to the solution of the modified sequence resulted in the formation of GQ, as the CD spectra showed, due to the removal of the reduction-sensitive p-nitrobenzyl group [322]. mG is one of the few natural base lesions was found so far to be repaired. O6-Alkylguanine-DNA alkyltransferase (AGT) is the repair enzyme that removes the promutagenic O6-alkylguanine adducts from duplex DNA in vitro and in vivo. Using mG-substituted A(htel-21) GQs, Hellman et al. [259] found that AGT repaired the mG adducts located within the folded GQ, and the rate of repair was comparable to those found with duplex DNAs under analogous conditions. Repair was dependent on the position of the adduct. In general, mG G3 located in the outermost (top and bottom) tetrads of a GQ stack exhibited more rapid-phase repair than did the adducts located in the inner tetrads.

_Hypoxanthine (I for Inosine)._ Enzymatic deamination of adenine nucleosides of RNA and DNA in vivo leads to hypoxanthine-ribosides, that is, to inosine or to 2'-deoxyinosine [299, 323]. Hypoxanthine, I (Figure 15, 6), is a G analog with a missing 2- amino group, and thus the G analog with a missing 2-amino group, and thus the G-to-I substitution at various G positions in the five TTAGG motifs produced the GIG motif with the strong hydrogen bond, and the reduced stacking interactions [321]. Bulky substituents linked to O₆ of G, such as the O₆-p-nitrobenzyl group, completely hindered the formation of GQ from the TBA oligodeoxynucleotide when substituted at G1. Adding Na₃S₂O₄ to the solution of the modified sequence resulted in the formation of GQ, as the CD spectra showed, due to the removal of the reduction-sensitive p-nitrobenzyl group [322]. mG is one of the few natural base lesions was found so far to be repaired. O6-Alkylguanine-DNA alkyltransferase (AGT) is the repair enzyme that removes the promutagenic O6-alkylguanine adducts from duplex DNA in vitro and in vivo. Using mG-substituted A(htel-21) GQs, Hellman et al. [259] found that AGT repaired the mG adducts located within the folded GQ, and the rate of repair was comparable to those found with duplex DNAs under analogous conditions. Repair was dependent on the position of the adduct. In general, mG G3 located in the outermost (top and bottom) tetrads of a GQ stack exhibited more rapid-phase repair than did the adducts located in the inner tetrads.

_Hypoxanthine (I for Inosine)._ Enzymatic deamination of adenine nucleosides of RNA and DNA in vivo leads to hypoxanthine-ribosides, that is, to inosine or to 2'-deoxyinosine [299, 323]. Hypoxanthine, I (Figure 15, 6), is a G analog with a missing 2-amino group, and thus the G analog with a missing 2-amino group, and thus the G-to-I substitution at various G positions in the five TTAGG motifs produced the GIG motif with the strong hydrogen bond, and the reduced stacking interactions [321]. Bulky substituents linked to O₆ of G, such as the O₆-p-nitrobenzyl group, completely hindered the formation of GQ from the TBA oligodeoxynucleotide when substituted at G1. Adding Na₃S₂O₄ to the solution of the modified sequence resulted in the formation of GQ, as the CD spectra showed, due to the removal of the reduction-sensitive p-nitrobenzyl group [322]. mG is one of the few natural base lesions was found so far to be repaired. O6-Alkylguanine-DNA alkyltransferase (AGT) is the repair enzyme that removes the promutagenic O6-alkylguanine adducts from duplex DNA in vitro and in vivo. Using mG-substituted A(htel-21) GQs, Hellman et al. [259] found that AGT repaired the mG adducts located within the folded GQ, and the rate of repair was comparable to those found with duplex DNAs under analogous conditions. Repair was dependent on the position of the adduct. In general, mG G3 located in the outermost (top and bottom) tetrads of a GQ stack exhibited more rapid-phase repair than did the adducts located in the inner tetrads.

_Hypoxanthine (I for Inosine)._ Enzymatic deamination of adenine nucleosides of RNA and DNA in vivo leads to hypoxanthine-ribosides, that is, to inosine or to 2'-deoxyinosine [299, 323]. Hypoxanthine, I (Figure 15, 6), is a G analog with a missing 2-amino group, and thus the G analog with a missing 2-amino group, and thus the G-to-I substitution at various G positions in the five TTAGG motifs produced the GIG motif with the strong hydrogen bond, and the reduced stacking interactions [321]. Bulky substituents linked to O₆ of G, such as the O₆-p-nitrobenzyl group, completely hindered the formation of GQ from the TBA oligodeoxynucleotide when substituted at G1. Adding Na₃S₂O₄ to the solution of the modified sequence resulted in the formation of GQ, as the CD spectra showed, due to the removal of the reduction-sensitive p-nitrobenzyl group [322]. mG is one of the few natural base lesions was found so far to be repaired. O6-Alkylguanine-DNA alkyltransferase (AGT) is the repair enzyme that removes the promutagenic O6-alkylguanine adducts from duplex DNA in vitro and in vivo. Using mG-substituted A(htel-21) GQs, Hellman et al. [259] found that AGT repaired the mG adducts located within the folded GQ, and the rate of repair was comparable to those found with duplex DNAs under analogous conditions. Repair was dependent on the position of the adduct. In general, mG G3 located in the outermost (top and bottom) tetrads of a GQ stack exhibited more rapid-phase repair than did the adducts located in the inner tetrads.

_Hypoxanthine (I for Inosine)._ Enzymatic deamination of adenine nucleosides of RNA and DNA in vivo leads to hypoxanthine-ribosides, that is, to inosine or to 2'-deoxyinosine [299, 323]. Hypoxanthine, I (Figure 15, 6), is a G analog with a missing 2-amino group, and thus the G analog with a missing 2-amino group, and thus the G-to-I substitution at various G positions in the five TTAGG motifs produced the GIG motif with the strong hydrogen bond, and the reduced stacking interactions [321]. Bulky substituents linked to O₆ of G, such as the O₆-p-nitrobenzyl group, completely hindered the formation of GQ from the TBA oligodeoxynucleotide when substituted at G1. Adding Na₃S₂O₄ to the solution of the modified sequence resulted in the formation of GQ, as the CD spectra showed, due to the removal of the reduction-sensitive p-nitrobenzyl group [322]. mG is one of the few natural base lesions was found so far to be repaired. O6-Alkylguanine-DNA alkyltransferase (AGT) is the repair enzyme that removes the promutagenic O6-alkylguanine adducts from duplex DNA in vitro and in vivo. Using mG-substituted A(htel-21) GQs, Hellman et al. [259] found that AGT repaired the mG adducts located within the folded GQ, and the rate of repair was comparable to those found with duplex DNAs under analogous conditions. Repair was dependent on the position of the adduct. In general, mG G3 located in the outermost (top and bottom) tetrads of a GQ stack exhibited more rapid-phase repair than did the adducts located in the inner tetrads.
only been studied via in silico methods [336, 337]. A complex quantum chemical approach showed a high degree of structural and energetic compatibility between the guanine and xanthine-based GQ models. The calculations established that hydrogen bonding made the greatest (~50%) contribution to the internal stability of the DNA GQs. Base stacking and ion coordination terms proved commensurable and accounted for the rest of the energetic contributions. Using two- or three-tetrad model systems, the guanine tetrad structures benefited from the high degree of H-bond cooperativity while the xanthine tetrad models were characterized by the more favorable stacking interactions [338]. Experimental proofs for these conclusions are not yet available.

**Cyclobutane Thymine Dimers.** Ultraviolet B (UVB) light irradiation induces the formation of cyclobutane pyrimidine dimers in nucleic acids in vivo. Under physiologically relevant solution conditions, the UVB light can prompt the formation of anti cyclobutane thymine dimers (Figure 16, 1) in GQs, as Smith et al. recently reported [339]. The anti refers to the position of the 5-methyl groups of the two thymines of the cyclobutane ring. Stereochemistry and yield of formation greatly depended on the topology of the GQ, the cations present, and the temperature of the solution. The two-tetrad basket-type model formed from various mutated analogs of the 26-mer A₃(htel-21)A₂, in which two-tetrad fold was first described for the (htel-21)T in K⁺ by Phan’s group [77], was found as the first GQ structure in which the thymine dimer could form in K⁺, but not in Na⁺. The T-dimer formed between the two edgewise loops. In this fold, the loop thymines are close enough for the UVB-induced cyclization reaction to occur. Other dynamic K⁺-topologies were also proposed as appropriate stereo structures for T-dimer formation. Thermal stability and conformational changes caused by the dimer formation were discussed by the authors [339].

**Clustered Lesions in GQs.** Ionizing radiation produces clustered lesions in the genetic material that results in distorted secondary structures in the double-stranded DNA. Clustered lesions have been found to be more difficult to repair, which were thus more harmful than the single, isolated ones and were detrimental to cells [340]. By definition, clustered DNA lesions form when two or more forms of damage appear in the double-stranded DNA strands within one to two helical turns, which involves 10–20 base pairs [341]. This type of lesions can also form in GQs. Sagi and coworkers [342] described for the first time the effect of clustered lesions on GQ structures, specifically clustered loop adenine (A/AP), tetrad guanine (G/AP), and A/AP + G/AP abasic sites (AP sites are among the most frequent natural lesions [299]). We have reformulated the definition described for duplex DNAs to be suitable for GQs as well: two or more lesions within a four-stranded, two- or three-tetrad core GQ unit consisting of up to 21 nucleotides in the core. For the experiments, the potential GQ-forming A(htel-21) and TA(htel-21) TTT oligodeoxynucleotides were synthesized with 36 permutations of double and triple tetrahydrofuranyl AP sites. In 100 mM K⁺ solution double loop A/AP sites destabilized the wild-type A(htel-21) GQ (Tm 66.1°C) by 6–8°C, whereas, when each of the three TTA loops contained an A/AP or one TTA loop was replaced by three A/AP sites, the wild type was even moderately stabilized, by 1–2°C. The unmodified GQ of TA(htel-21)TT (Tm: 62.0°C) was extensively destabilized by single tetrad G/APs, by 15 to 36°C; the latter figure belonged to the middle tetrad substitutions of the three-tetrad GQ. Depending on its position, the single A/APs diminished or intensified the damaging effect of G/APs. The largest negative effects were observed when the G/AP was in the middle tetrad: only partially folded structures were formed at 37°C. Half of the 21 variants of double and triple clustered G/AP sites led to more or
less unfolded structures at 37°C; others had $T_m$ between 31 and 37°C, meaning partial folding only. Three of those oligodeoxynucleotides containing one G/AP in the middle tetrad remained unstructured. The single to triple A/APs in the loops increased the population of parallel strands in their structures, thus shifting the antiparallel A(htel-21) GQs into hybrid or parallel forms. Single G/APs and thus also the G/AP plus A/AP combinations, however, inhibited the formation of parallel strands and these GQs adopted antiparallel topologies. CD spectra matched those of the Na+-stabilized htel GQs. The in vitro results may suggest that formation of clustered lesions in the chromosome-capping structure, in the Shelterin complex, can induce unfolding of existing GQ structures, which, in turn, can lead to telomere shortening [342].

5.3. Epigenetic Modifications

5-Methyl- and 5-Hydroxymethylcytosines ($m^5$C and $hm^5$C).

Epigenetic 5-methylation of cytosine is among the processes that control the gene expression in human cells, which has also associations with the development of cancer and other diseases [343–345]. Methylation usually occurs at CpG sites and 70–80% of the CpG sites in mammalian cells are methylated. $m^5$C accounts for ~1% of all DNA bases in the human genome [344, 346]. Methylation is a reversible enzymatic process [347]. Demethylation of $m^5$C (Figure 16, 2) follows stepwise through 5-hydroxymethylcytosine ($hm^5$C) (Figure 16, 3), to 5-formylcytosine and finally to 5-carboxycytosine, and this base is then excised by the subsequent repair processes. Interestingly, all these bases have been found to occur in mammalian embryonic stem (ES) cells [348], indicating that the modified cytosines are stable and may themselves have some role in gene regulation. Their abundance in cells is very low, ~0.5%, ~0.002%, and 0.0003%, respectively, for the three oxidized derivatives of $m^5$C, in mouse ES cells [349]. Their level also depends on age, at least in the brain [350]. Regarding the presence of $m^5$C in CpG sites of GQs, the PI promoter sequence of bcl-2 oncogene, the CGGCGCGCGGTTATGCTCGGAGC was first examined, where the cytosines at CI, C5, C7, and C21 positions were methylated. The methylations induced the formation of GQ in the PI promoter by greatly increasing the thermal stability of the structure. Methylation also protected the GQ from unfolding by the complementary strand. The proposed mechanism explained the transcription inactivation by hypermethylation that induces formation of GQ from duplex in the promoter and the gene activation by hypomethylation [351]. The 5-hydroxymethylcytosines ($hm^5$C) are also involved in the regulation of gene expression of certain genes. The studies described by Molnar et al. [352] characterized the effects of $hm^5$C-modified CpG islands on the GQ (and the i-motif) structures. Single $hm^5$C was incorporated into the G-rich (and C-rich) sequences from the vascular endothelial growth factor (VEGF) promoter. The results indicated that two of the three $hm^5$C-containing loops showed a significant decrease in thermal stability of the GQ (on the contrary, thermal stability of the i-motif increased somewhat by the $hm^5$C) [352]. Another study was carried out with the G-rich and C-rich strand models of the C9orf72 repeat, the (GGGGCC)$_8$ and (GGCCCC)$_8$ DNA oligonucleotides, to investigate the effect of epigenetic signals on the macromolecular structure. Each C was replaced by $m^5$C or $hm^5$C at each CpG motive. The G-rich strands formed G-tetrad- and various mixed G + C tetrad-containing unusually long intramolecular chair-type antiparallel GQs. GQs of the C-rich strands contained C tetrads and also mixed G + C tetrads while no i-motif was observed. The $m^5$C moderately increased, by 1.2 C, while $hm^5$C considerably decreased, by 4.7 C, the thermal stability of the unmodified intramolecular GQ of the G-rich strand in K$^+$ [353].

8-Oxoguanine ($o^8$G). ROS, the reactive forms of oxygen, are free radicals that have unpaired electrons in their outer orbits. ROS can cause significant damage in cells, where DNA is one of the major targets. Recent studies, however, also demonstrated that ROS have normal physiological functions as well. For instance, $o^8$G (Figure 15, 1), one of the initial main damage products of ROS, has recently been described by Fleming et al. [354] as it is also an epigenetic modification in gene promoters that regulate the transcription via the repair process of the base excision repair (BER). Gene expression was found to occur when $o^8$G is formed in the G-rich, potentially GQ-forming DNA sequences. The damage initiated the action of the BER by the 8-oxo-2'-deoxyguanosine DNA glycosylase (hOGG1), whose reaction yields an abasic (AP) site. The AP site enabled unwinding (melting) of the duplex DNA that facilitated the conversion of the G-rich strand to a GQ fold. These reactions finally resulted in the activation of the VEGF or NTHLI genes. The authors identified 61 human DNA repair genes that might be activated by this mechanism [354]. Later, Fleming and Burrows have confirmed these findings and also compared the epigenetic pathway of $o^8$G with that of another base analog, the 5-methylcytosine [355].

6. Concluding Remarks

This study illustrates the established, widespread use of synthetic nucleotide analogs in the structural studies of the most researched noncanonical nucleic acid structures, the GQs. Application of the discussed, known analogs and introduction of new derivatives is not going to stop; thus we can expect further growth of the field. Significant advances can be anticipated in the elucidation of the effect of numerous, unexplored single and clustered natural base lesions on the structures and repair of GQs. Recent findings of epigenetics on the role of the formation of some oxidative DNA base lesions followed by conformation switching of the G-rich duplex DNAs to GQs suggest the opening of a new area in the gene expression research.

Abbreviations

GQ: G-quadruplex

$T_m$: Half-way temperature of the thermal unfolding profile of the GQs

htel: Human telomere repeat sequences
TBA: Thrombin-binding aptamer
CD: Circular dichroism.

Additional Points
In most cases, DNA GQs are discussed here, and therefore the “d” is omitted from the sequence; for instance, d(TGGGAG) is denoted as TGGGAG, or the htel DNA, such as with small letters, like ugggu.

Conflicts of Interest
The author declares that there are no conflicts of interest.

References


